

# **Gestaltung unternehmensübergreifender Business-Intelligence-Netzwerke**

Von der Mercator School of Management, Fakultät für Betriebswirtschaftslehre, der

Universität Duisburg-Essen

zur Erlangung des akademischen Grades

eines Doktors der Wirtschaftswissenschaft (Dr. rer. oec.)

genehmigte Dissertation

von

Jens Kaufmann

aus

Rinteln

Referent: Prof. Dr. Peter Chamoni

Korreferent: Prof. Dr. Alf Kimms

Tag der mündlichen Prüfung: 16.12.2015

**Inhaltsverzeichnis**

<b>Abbildungsverzeichnis .....</b>	<b>V</b>
------------------------------------	----------

<b>Tabellenverzeichnis .....</b>	<b>VIII</b>
----------------------------------	-------------

<b>Abkürzungsverzeichnis .....</b>	<b>X</b>
------------------------------------	----------

<b>1 Einleitung .....</b>	<b>1</b>
---------------------------	----------

1.1 Motivation und Zielsetzung der Arbeit .....	1
---	---

1.2 Wissenschaftstheoretische Positionierung .....	3
--	---

1.2.1 Einordnung der gestaltungsorientierten Wirtschaftsinformatik .....	5
--	---

1.2.2 Forschungsmethoden der Wirtschaftsinformatik .....	8
--	---

1.3 Aufbau der Arbeit.....	11
----------------------------	----

<b>2 Zusammenarbeit von Unternehmen.....</b>	<b>13</b>
--	-----------

2.1 Definition und Motivation.....	14
------------------------------------	----

2.2 Kriterien für Unternehmenskooperationen.....	17
--	----

2.2.1 Ansätze in der Literatur .....	18
--------------------------------------	----

2.2.2 Kooperationsrichtung.....	21
---------------------------------	----

2.2.3 Kooperationsform.....	23
-----------------------------	----

2.2.4 Kooperationsintensität.....	24
-----------------------------------	----

2.3 Ausgewählte Konzepte unternehmensübergreifender Kooperation.....	25
--	----

2.3.1 Strategische Allianzen.....	26
-----------------------------------	----

2.3.2 Supply Chain Management .....	27
-------------------------------------	----

2.4 Fallbeispiel zur Kooperation in Analyseprozessen: Auswertung der gemeinsamen Produktentwicklung in Zulieferernetzwerken .....	28
--	----

<b>3 Entscheidungsunterstützung in Unternehmen .....</b>	<b>34</b>
--	-----------

3.1 Business Intelligence .....	34
---------------------------------	----

3.1.1 Definition und Einordnung in den Kontext von Decision- Support-Systemen .....	35
--	----

3.1.2 Data Warehousing .....	39
------------------------------	----

3.1.3 Multidimensionale Datenhaltung und Analyse .....	42
--	----

3.2 Gruppenentscheidungsunterstützung.....	49
--	----

3.2.1 Gruppenentscheidungsprozesse .....	50
--	----

3.2.2 Systemische Unterstützung .....	52
---------------------------------------	----

3.2.3 Veränderungen durch den Einfluss sozialer Medien.....	56
---	----

3.3 Collaborative Business Intelligence .....	57
---	----

3.3.1 Literaturüberblick.....	58
-------------------------------	----

3.3.2 Aufbau eines Ordnungsrahmens .....	65
--	----

3.3.3 Eigenschaften unterschiedlicher Formen .....	70
--	----

3.3.4 Anwendung auf unternehmensübergreifende Kooperation in Analyseprozessen .....	76
--	----

<b>4</b>	<b>Konzept eines unternehmensübergreifenden Business-Intelligence-Netzwerks .....</b>	<b>78</b>
4.1	Cross-Company Business Intelligence .....	79
4.2	Paralleles und verteiltes Arbeiten im Data-Warehouse-Kontext .....	84
4.2.1	Verteilte Data-Warehouse-Systeme .....	86
4.2.2	Peer-to-peer-basiertes Online Analytical Processing .....	92
4.3	Harmonisierung und Interpretierbarkeit multidimensionaler Datenstrukturen .....	96
4.3.1	Konflikte unterschiedlicher multidimensionaler Datenstrukturen ....	97
4.3.2	Grundlegende Lösungsansätze für verteilte Datenbasen.....	101
4.3.3	Data-Warehouse-spezifische und Ontologie-basierte Ansätze .....	103
4.3.4	Ansätze mit besonderer Berücksichtigung unabhängiger Netzteilnehmer .....	108
4.4	Sicherheit und Vertrauen in Business-Intelligence-Netzwerken.....	112
4.5	CroCoBIN: Ein einfaches Referenzmodell für Peer-to-peer-Business-Intelligence-Netzwerke .....	114
4.5.1	Abgrenzung des Referenzmodellbegriffs.....	115
4.5.2	Aufbau des Referenzmodells .....	118
4.5.3	Gesamtmodell mit Einordnung bestehender Ansätze und Ableitung notwendiger Entwicklungen.....	130
<b>5</b>	<b>Sicherstellung der Nutzbarkeit von Business-Intelligence-Netzwerken .....</b>	<b>133</b>
5.1	Exkurs: Daten- und Informationsqualität .....	135
5.2	Mess- und Einflussgrößen in Business-Intelligence-Netzwerken.....	138
5.2.1	Struktur und Beziehungen der Netzwerkelemente.....	138
5.2.2	Aufbau und Eigenschaften von Netzwerken .....	143
5.2.3	Methodik zur Erarbeitung von Messgrößen.....	148
5.3	Definition der Messgrößen .....	151
5.3.1	Mapping-Qualität .....	151
5.3.2	Pfad-Qualität .....	156
5.3.3	Routing-Qualität.....	160
5.3.4	Netz-Qualität .....	162
5.4	Einflussfaktoren auf die Güte des Netzes.....	165
<b>6</b>	<b>Bestimmung sinnvoller Konfigurationen von Business-Intelligence-Netzwerken .....</b>	<b>168</b>
6.1	Strukturierung und Abgrenzung von Simulationsverfahren .....	168
6.2	Modell des Fallbeispiels.....	173
6.3	CroCoSIM: Prototypische Implementierung einer Simulation für Peer-to-peer-Business-Intelligence-Netzwerke.....	175
6.3.1	Systemarchitektur und Implementierung .....	176
6.3.2	Umfang und Konfigurationsmöglichkeiten.....	179
6.3.3	Auswertungsmöglichkeiten .....	184

---

6.4	Durchführung der Simulation und Interpretation der Ergebnisse .....	185
6.4.1	Gestaltung der Mapping-Qualität.....	187
6.4.2	Auswahl geeigneter Kombinationsverfahren zur Bestimmung der Routing-Qualität.....	192
6.4.3	Betrachtung unterschiedlicher Netzstrukturen, Rahmenbedingungen und Strategien zur Nachbarschaftswahl .....	199
6.4.4	Aussagekraft und Prognosefähigkeit der Gütemaße .....	219
6.4.5	Abschließende Betrachtungen der Ergebnisse und des Fallbeispiels.....	225
<b>7</b>	<b>Resümee und Ausblick.....</b>	<b>227</b>
7.1	Zusammenfassung der Arbeit.....	227
7.2	Schlussfolgerungen aus den Untersuchungen und Implikationen auf die weitere Forschung .....	229
	<b>Literaturverzeichnis .....</b>	<b>232</b>
	<b>Anhangsverzeichnis .....</b>	<b>266</b>

**Abbildungsverzeichnis**

Abb. 1: Generelle Struktur eines Informationssystems .....	4
Abb. 2: Begründungsinstanzen der Bewertung von Forschungsergebnissen .....	5
Abb. 3: Schematische Darstellung zentraler erkenntnistheoretischer Positionen .....	6
Abb. 4: Empirisch gestütztes Methodenprofil der Wirtschaftsinformatik .....	9
Abb. 5: Design Science Research Methodology .....	10
Abb. 6: Modell einer Wertkette .....	13
Abb. 7: Synthese morphologischer Kästen und Kriteriumsbeschreibungen für unternehmensübergreifende Kooperation .....	19
Abb. 8: Regelkreis der Unternehmenssteuerung .....	30
Abb. 9: Ausprägungen der Kooperationskriterien im Fallbeispiel .....	33
Abb. 10: DSS-Zeitlinie .....	37
Abb. 11: BI-Evolution .....	37
Abb. 12: Einordnung unterschiedlicher Facetten von Business Intelligence .....	38
Abb. 13: Data-Warehouse-Architektur .....	41
Abb. 14: Schematische Darstellung eines OLAP-Würfels .....	43
Abb. 15: Beispielhafte Darstellung einer 'Geografie'-Hierarchie .....	44
Abb. 16: Beispielhafte Modellierung einer Dimension 'Geografie' mit ADAPT als Bestandteil eines Würfels .....	45
Abb. 17: Beispiel eines Star-Schemas .....	48
Abb. 18: Prozess des Knowledge Discovery in Databases .....	49
Abb. 19: Grundmodell des Entscheidungsprozesses .....	51
Abb. 20: Dimensionen der Forschung zu Gruppenentscheidungsunterstützungssystemen .....	54
Abb. 21: Klassifikationen von Systemen zur Gruppenzusammenarbeit .....	55
Abb. 22: Phasen der Reviewforschung .....	59
Abb. 23: Einordnung der Literaturrecherche in die Taxonomie von COOPER .....	60
Abb. 24: Anzahl Publikationen zu CBI nach Jahren .....	64
Abb. 25: Ordnungsrahmen für Collaborative Business Intelligence .....	68
Abb. 26: BIN-Architektur .....	81
Abb. 27: Unterschiedliche Architekturen für paralleles Rechnen .....	85
Abb. 28: Unterschiedliche DWH-Architekturen (1) .....	87
Abb. 29: Unterschiedliche DWH-Architekturen (2) .....	88

Abb. 30: Schematischer Aufbau eines P2P-OLAP-Netzwerks mit unterschiedlicher Knoten-Definition .....	93
Abb. 31: Beispielhafte Instanz einer heterogenen Hierarchie .....	98
Abb. 32: Klassifikation möglicher Konflikte bei multidimensionalen Datenstrukturen .....	99
Abb. 33: Kennzahlen- und Kontenschema im Vergleich an einem Beispiel .....	100
Abb. 34: Umordnung von Elementen einer ‚Produkt‘-Dimension.....	110
Abb. 35: Aufteilung von Elementen einer ‚Produkt‘-Dimension.....	110
Abb. 36: Grundstruktur eines CroCoBIN.....	118
Abb. 37: Grundstruktur und Komponenten eines CroCoBIN .....	120
Abb. 38: OSI-Schichtenmodell.....	122
Abb. 39: Peerstruktur in einem CroCoBIN .....	125
Abb. 40: CroCoBIN-Referenzmodell (Gesamtdarstellung) .....	130
Abb. 41: Graph eines einfachen CroCoBIN .....	139
Abb. 42: Graph eines einfachen CroCoBIN in angepasster Notation .....	140
Abb. 43: Schematischer Zusammenhang von Netzwerkelementen, Messgrößen und Einflussfaktoren zur Qualitätsmessung in einem CroCoBIN.....	143
Abb. 44: Vergleich unterschiedlicher Netzwerk-Topologien.....	144
Abb. 45: Vergleich eines Zufallsnetzwerks mit einem Skalenfreien Netzwerk..	147
Abb. 46: Werte von Poisson-Verteilungen und Potenzgesetzen abhängig von $k$ .....	147
Abb. 47: Zwei TMQ-identische Abdeckungen eines Master-Schemas mit unterschiedlicher $MQ_s/MQ_c$ .....	154
Abb. 48: Winkel zwischen Vektoren unterschiedlicher Abdeckungen gleicher TMQ und dem Vektor des Master-Schemas .....	155
Abb. 49: Darstellung eines einfachen Pfades zur Ermittlung der TPQ .....	157
Abb. 50: Plots unterschiedlicher t-Normen .....	158
Abb. 51: Alternative Routing-Möglichkeiten in einem 5-elementigen Netzwerk.....	160
Abb. 52: Vorgehensmodell zur Erstellung einer Simulation.....	170
Abb. 53: Umfang des Modells bezogen auf das CroCoBIN-Referenzmodell.....	173
Abb. 54: ADAPT-Darstellung des Würfels zum Fallbeispiel .....	174
Abb. 55: Vereinfachtes UML-Klassendiagramm von CroCoSIM .....	177
Abb. 56: Grafische Netz- und Routing-Darstellungen in CroCoSIM .....	184

Abb. 57: Diagramm-Darstellungen in CroCoSIM .....	185
Abb. 58: TMQ-Werte in Abhängigkeit von der MQ mit Gewichtung nach Ergebnissen der multiplen Regression .....	189
Abb. 59: TMQ-Werte in Abhängigkeit von den Teilqualitäten der MQ .....	189
Abb. 60: TNQ in Abhängigkeit von der NQ sowie NQ und Netzwerkgröße unter Verwendung von $T^H_1$ .....	194
Abb. 61: 3D-Plot der t-Norm $T^Y_{1,25}$ .....	198
Abb. 62: Wertebereiche der durchschnittlichen NQ in Abhängigkeit von einzelnen Parametern .....	200
Abb. 63: Anteil Netzwerke mit gegebener NQ nach Größenklassen .....	201
Abb. 64: NQ unterschiedlicher Netzwerke in Abhängigkeit von der Gradanzahl .....	202
Abb. 65: NQ unterschiedlicher Netzwerke in Abhängigkeit von der PMQG .....	203
Abb. 66: NQ, AvgPL und CC in Abhängigkeit von Netztyp und -größe .....	204
Abb. 67: NQ skalenfreier Netzwerke in Abhängigkeit von $\mu$ und Knotensortierung .....	205
Abb. 68: NQ in Abhängigkeit von Netztyp, Netzgröße und ‚selfish‘- Parameter .....	205
Abb. 69: NQ von Zufallsnetzwerken in Abhängigkeit von PMQG und ‚selfish‘-Parameter .....	206
Abb. 70: TNQ in Abhängigkeit von der NQ für unterschiedliche Netzwerke .....	208
Abb. 71: Netzwerkanteile und NQ-Werte für Zufallsnetzwerke .....	213
Abb. 72: NQ von Zufallsnetzwerken in Abhängigkeit vom Spread .....	213
Abb. 73: Durchschnittliche Standardabweichung der RQ in Zufallsnetzwerken in Abhängigkeit vom Spread .....	214
Abb. 74: NQ in Abhängigkeit von PMQG, Spread und Netzwerkgröße .....	215
Abb. 75: NQ von Zufallsnetzwerken in Abhängigkeit von der Dichte .....	216
Abb. 76: TMQ und geschätzte TMQ in Abhängigkeit von der MQ .....	221
Abb. 77: 3D-Plot der TNQ von Zufallsnetzwerken in Abhängigkeit von Netzgröße und NQ .....	222
Abb. 78: TNQ von Zufallsnetzwerken in Abhängigkeit von NQ und Gradanzahl für unterschiedliche Netzgrößen .....	223



## Tabellenverzeichnis

Tab. 1: Ausgewählte Methoden der gestaltungsorientierten Wirtschaftsinformatik.....	9
Tab. 2: Behandlung von Kriterien zur Unterscheidung von Kooperationen in unterschiedlichen Publikationen.....	19
Tab. 3: Weitere Kriterien zur Unterscheidung von Kooperationsarten (weniger als drei Nennungen) .....	20
Tab. 4: Zuordnung von Datenbanken und Verlagsseiten zu Publikationen .....	62
Tab. 5: Anzahl Treffer und verwendeter Beiträge der Literatursuche.....	63
Tab. 6: Dimensionen und Ausprägungen des Ordnungsrahmens für Collaborative Business Intelligence .....	67
Tab. 7: Zuordnung der Publikationen zum Ordnungsrahmen für Collaborative Business Intelligence .....	68
Tab. 8: Anzahl der Beiträge in der Zuordnung zur TOFE-Klassifikation.....	70
Tab. 9: Zuordnung der Modell-Konzepte des SP2P-Referenzmodells zu den Komponenten von CroCoBIN .....	121
Tab. 10: Zuordnung vorgestellter Ansätze zu Komponenten und zentralen Aspekten des CroCoBIN-Referenzmodells.....	131
Tab. 11: Konstruktionsübersicht unterschiedlicher t-Normen .....	159
Tab. 12: Bewertungsunterschiede bei Einsatz unterschiedlicher t-Normen.....	159
Tab. 13: Parameter zur Definition der Schemagenerierung .....	180
Tab. 14: Parameter zur Gestaltung des Netzaufbaus .....	181
Tab. 15: Parameter zur Beeinflussung der Nachbarschaftswahl .....	183
Tab. 16: Übersicht der durchgeführten Simulationen .....	186
Tab. 17: Korrelationskoeffizienten unterschiedlicher MQ-Varianten und TMQ .....	191
Tab. 18: Auswertung unterschiedlicher t-Normen .....	196
Tab. 19: Parameterkonfigurationen der ersten Netzwerk-Simulationsreihe.....	199
Tab. 20: Parameterkonfigurationen der zweiten Netzwerk-Simulationsreihe.....	207
Tab. 21: Durchschnittliche NQ und TNQ für unterschiedliche Netzwerke .....	208
Tab. 22: p-Werte der t-Tests zur Überprüfung einer Abweichung der durchschnittlichen TNQ bei unterschiedlicher Konfiguration .....	211
Tab. 23: Einfluss von Variablen und Interaktionstermen auf die NQ .....	217
Tab. 24: Einfluss von Variablen und Interaktionstermen auf die TNQ.....	218

---

Tab. 25: Bestimmtheitsmaße Gradanzahl-abhängiger und -unabhängiger Regressionsanalysen von TNQ und NQ.....	223
Tab. 26: Zusammenfassung der Ergebnisse der einzelnen Simulationen.....	226

**Abkürzungsverzeichnis**

ADAPT	Application Design for Analytical Processing Technologies
AMZ	Netzwerk Automobilzulieferer Sachsen
AvgPL	Average Path Length
BAV	Both-As-View
BI	Business Intelligence
BIN	Business Intelligence Network
BSC	Balanced Scorecard
CBI	Collaborative Business Intelligence
CC	Cluster Coefficient
CCBI	Cross-Company Business Intelligence
CDM	Collaborative Decision Making
CG	Computing Grid
CPFR	Collaborative Planning, Forecasting & Replenishment
CRM	Customer Relationship Management
DQ	Datenqualität
DSS	Decision-Support-System
DWH	Data Warehouse
EDI	Electronic Data Interchange
EIS	Executive Information Systems
ERP	Enterprise Resource Planning
ESS	Executive Support Systems
ETL	Extract, Transform, Load
F&E	Forschung und Entwicklung
GAV	Global-As-View
GDSS	Group Decision Support Systems
GI-FB WI	Fachbereich Wirtschaftsinformatik der Gesellschaft für Informatik
GLAV	Global-Local-As-View
I/O	Input / Output
IC	Internal Communication
IQ	Informationsqualität
IR	Information Retrieval
IS	Information Systems
ISO	International Organization for Standardization

---

IT	Informationstechnologie
JIT	Just In Time
JV	Joint Venture
KDD	Knowledge Discovery in Databases
KMU	Kleine und Mittlere Unternehmen
LAV	Local-As-View
M&A	Mergers & Acquisitions
MIS	Management Information Systems
MQ	Mapping-Qualität
MQ <sub>b</sub>	Mapping-Qualität (Balance)
MQ <sub>c</sub>	Mapping-Qualität (Consistency)
MQ <sub>s</sub>	Mapping-Qualität (Soundness)
MRA	Modified Random Algorithm
$\mathcal{MD}$	Multidimensional Data Model
NGO	Nichtregierungsorganisation
NQ	Netz-Qualität
ODS	Operational Data Storage
OLAP	On-Line Analytical Processing
OR	Operations Research
OSI	Open Systems Interconnection
OSN	Online Social Network
P2P	Peer-to-peer
PA	Partnership in Analysis
PD	Partnership in Data
PDCA	Plan, Do, Check, Act
PDMS	Peer-Data-Management-System
PMQ	Prognostizierte Mapping-Qualität
PMQG	Prognostizierte-Mapping-Qualität-Güte
PPS	Produktionsplanung und -steuerung
PQ	Pfad-Qualität
RDF	Resource Description Framework
RQ	Routing-Qualität
SA	Sensitivitätsanalyse
SBI	Social Business Intelligence
SCM	Supply Chain Management

---

SD	Standardabweichung
SP2P	Semantic Peer-to-peer
TMQ	Tatsächliche Mapping-Qualität
TNQ	Tatsächliche Netz-Qualität
TOFE	Technisch, Organisatorisch, Fachlich, Ökonomisch
TPQ	Tatsächliche Pfad-Qualität
TRQ	Tatsächliche Routing-Qualität
UML	Unified Modeling Language
VMI	Vendor-Managed Inventory
W3C	World Wide Web Consortium
WI	Wirtschaftsinformatik
WKWI	Wissenschaftliche Kommission Wirtschaftsinformatik
WWW	World Wide Web
XML	Extensible Markup Language

## 1 Einleitung

Unternehmen aller Branchen agieren heute in einem Umfeld, das durch umfangreiche Vernetzung und ein hohes technologisches Niveau geprägt ist. Der durchgängige Einsatz von Informations- und Kommunikationstechnik führt dabei sowohl auf inner- als auch auf überbetrieblicher Ebene zu der Notwendigkeit, Kooperationen einzugehen, zu verstärken und in hoher Geschwindigkeit abzuwickeln.<sup>1</sup> Die stetig sinkenden Latenzzeiten zur Abstimmung verlagern zudem in Kombination mit sinkenden Kosten für Kommunikation, Informationsbeschaffung, -verarbeitung und -verteilung die Entscheidungsbedürfnisse und -befugnisse von zentralen zu dezentralen Stellen.<sup>2</sup> Der Austausch von Daten und Informationen nimmt daher stetig zu und stellt die Beteiligten vor die Aufgabe, nicht nur eigene, sondern auch fremde Quellen zu erfassen, aufzubereiten und die gewonnenen Erkenntnisse managementfreundlich darzustellen sowie es Analysten zu ermöglichen, weitere Einsichten zu gewinnen. *Business Intelligence* (BI)<sup>3</sup> ist dabei seit über 20 Jahren der dominierende Begriff für Konzepte, Methoden und Systeme zur Entscheidungsunterstützung in Unternehmen.<sup>4</sup>

### 1.1 Motivation und Zielsetzung der Arbeit

Sowohl Entscheidungsunterstützung im Allgemeinen als auch Business Intelligence im Besonderen sind in Wissenschaft und Praxis ausführlich betrachtet und diskutiert worden. Ihre ursprüngliche Fokussierung auf die innerbetriebliche Welt hat jedoch dazu geführt, dass ihre Evolution zur Unterstützung unternehmensübergreifender Zusammenarbeit nicht so schnell verlaufen ist, wie die betriebliche Praxis Kooperationen ausgebaut hat. Im Rahmen der gemeinsamen Wertschöpfung existiert eine durch Begriffe wie *Supply Chain Management* (SCM) geprägte Welt der Zusammenarbeit zwischen Unternehmen, ihren Lieferanten und Abneh-

---

<sup>1</sup> Vgl. Kagermann (2014): 606f., 613.

<sup>2</sup> Vgl. Bloom et al. (2014): 2859, 2877.

<sup>3</sup> Die vorliegende Arbeit verwendet – neben den vorgegebenen – folgende Gestaltungsrichtlinien: *Kursiv* gesetzt werden zentrale Fachbegriffe bei ihrem ersten Auftreten sowie Emphasen im Sinne sprachlicher Hervorhebungen oder Betonungen. In einfache Anführungszeichen (‘ ’) gesetzte Ausdrücke sind einzelne, nicht wörtliche Wiedergaben von Ausführungen Dritter oder Begriffe, die umgangssprachlich ‚in Anführungszeichen‘ gesetzt werden. Doppelte Anführungszeichen („ “) werden ausschließlich für wörtliche Zitate genutzt. Erläuterungen, Abkürzungen, Übersetzungen und Quellhinweise werden mit runden Klammern dargestellt, sprachliche Einschübe mit einem Halbgeviertstrich (–) abgetrennt. Bei der Angabe mehrerer Autoren in Quellenangaben sind diese alphabetisch geordnet.

<sup>4</sup> Vgl. Watson/Wixom (2007): 96.

mern. Unternehmen gleicher Produktions- oder Leistungsstufe kooperieren als Netzwerke und Allianzen z. B. im Flugverkehr oder als Einkaufs-, Vertriebs- und Entwicklungsgemeinschaften.

Für die kooperative Arbeit im Bereich BI existieren hingegen kaum ähnlich prominente und erforschte Lösungen. Die Entwicklung von Systemen über Organisationsgrenzen hinweg und eine umfassende Kooperation in diesem Bereich sind daher als ein zentraler Innovationsstrang für BI von der deutschen Wirtschaftsinformatik identifiziert worden.<sup>5</sup> Unter dem Begriff *Collaborative Business Intelligence* (CBI) beginnen Wissenschaft und Praxis seit wenigen Jahren, Konzepte und Systeme zu entwickeln, die die innerbetriebliche Kommunikation und Kooperation in bestehenden BI-Landschaften, gelegentlich auch die überbetriebliche Vernetzung zur gemeinsamen Datenauswertung, ermöglichen oder verbessern sollen.<sup>6</sup> Als Grundkonstellation werden dabei zentral oder dezentral ausgerichtete Netzwerke betrachtet. Zentrale Lösungen mit ‚federführenden‘ Netzwerkteilnehmern, die dem bisherigen Ansatz von unternehmensspezifischen Systemen nahe kommen, stehen dabei im Fokus. *Peer-to-peer*(P2P)-basierte Ansätze – als vollständig dezentrale Konzepte – werden trotz der Popularität von P2P-Netzwerken im letzten Jahrzehnt und der damit einhergehenden wissenschaftlichen Fundierung kaum in Betracht gezogen.<sup>7</sup> Dabei bieten P2P-Ansätze einige wertvolle Grundeigenschaften im Kontext komplexer Unternehmensnetzwerke: Sie sind flexibel in Bezug auf Größe und Teilnahme(dauer) und ermöglichen einen maximalen Autonomiebehalt.

---

<sup>5</sup> Vgl. Baars et al. (2014): 16.

<sup>6</sup> MERTENS weist darauf hin, dass häufig eine wortklangähnliche und nicht sinngemäße Übertragung von Anglizismen in die deutsche Sprache erfolgt. Er nennt als typisches Beispiel die Übertragung von ‚collaboration‘ in ‚Kollaboration‘ und spricht von einer ‚Scheinübersetzung‘ (vgl. Mertens (2006): 36). Der DUDEN definiert Kollaboration als „gegen die Interessen des eigenen Landes gerichtete Zusammenarbeit mit dem Kriegsgegner, mit der Besatzungsmacht“ (Duden (2015b)). In der englischen Sprache ist diese Bedeutung von ‚collaboration‘ zwar auch bekannt, wesentlich häufiger – insbesondere in Zusammenhang mit der gemeinschaftlichen Arbeit von Unternehmen oder Personen – wird jedoch die Bedeutung der deutschen Begriffe Zusammenarbeit oder Kooperation darunter verstanden (vgl. Oxford Dictionaries (2015); Duden (2015c)). In dieser Arbeit werden daher grundsätzlich die Begriffe ‚Kollaboration‘ und ‚kollaborativ‘ nicht verwendet. Die Begriffe ‚collaboration‘ und Kooperation (oder Zusammenarbeit) werden hingegen synonym betrachtet, wobei der englische Begriff in der Regel nur in entsprechenden Namen, Zitaten oder englischen Passagen zum Tragen kommt (vgl. hierzu auch die Ausführungen in Abschnitt 2.1). Es sei darauf hingewiesen, dass andere Autoren dem in dieser Arbeit gewählten Weg nicht folgen. So leitet LEIMEISTER aus den Ausführungen der Fachliteratur ab, dass Kollaboration eine stärker gruppenzielorientierte und technisch unterstützte Form der Zusammenarbeit darstellt und verwendet ‚Kollaboration‘ als Oberbegriff für Kooperation, Kommunikation und Koordination (vgl. Leimeister (2014): 7–9).

<sup>7</sup> Für eine ausführlichere Diskussion der Thematik vgl. Abschnitt 4.1.

Die vorliegende Arbeit diskutiert daher den Einsatz verschiedener, insbesondere jedoch P2P-basierter Netzwerke für unternehmensübergreifende Kooperation im Bereich BI. Die Zielsetzung der Arbeit lässt sich dabei nach Erkenntnis- und Gestaltungsziel aufspalten.<sup>8</sup>

Das Erkenntnisziel besteht in der Prüfung der Eignung von P2P-BI-Netzwerken für die unternehmensübergreifende analytische Kooperation von Unternehmen. Neben der allgemeinen Betrachtung ist zu bewerten, wie sich unterschiedliche Rahmenbedingungen und Netzwerkkonfigurationen auf die Nutzungsfähigkeit unter Bedingungen eines betrieblichen Kontextes auswirken. Dabei sind bestehende betriebswirtschaftliche Erkenntnisse, z. B. über das größenabhängige Verhalten von Netzwerken, bzw. deren Teilnehmern, zu berücksichtigen.<sup>9</sup>

Zu diesem Zweck wird ein Referenzmodell eines P2P-BI-Netzwerks entworfen. Darauf aufbauend werden unter Berücksichtigung bestehender Ansätze Gütekriterien für derartig gestaltete Netzwerke entwickelt. Mit Hilfe einer Vielzahl prototypisch instanzierter Modelle werden unterschiedliche Netzwerkkonfigurationen auf ihre Güte untersucht und Empfehlungen zur Netzgestaltung abgeleitet. Kriterien-, Modell- und Instanz-Entwicklung stellen das Gestaltungsziel dieser Arbeit dar.<sup>10</sup>

## 1.2 Wissenschaftstheoretische Positionierung

Die *Wirtschaftsinformatik* (WI) hat sich seit ihren Anfängen als eigenständige Wissenschaft positioniert und formuliert diesen Anspruch über einen klar beschriebenen Gegenstandsbereich.<sup>11</sup> Im deutschsprachigen Raum ist dieser im *Profil der Wirtschaftsinformatik* beschrieben: „Gegenstand der Wirtschaftsinformatik sind Informationssysteme (IS) in Wirtschaft, Verwaltung und privatem Bereich.“<sup>12</sup>

---

<sup>8</sup> Ein *Erkenntnisziel* fokussiert das Verständnis gegebener Sachverhalte, ein *Gestaltungsziel* die Schaffung neuer Sachverhalte. Vgl. für eine ausführliche Darstellung und Aufgliederung Becker et al. (2003): 11f.

<sup>9</sup> Dies betrifft bspw. den Effekt der ‚kritischen Masse‘, also einer Mindestgröße zur wirtschaftlich sinnvollen Nutzung von Netzwerken (vgl. dazu Katz/Shapiro (1994): 93–97; sowie die Ausführungen in Abschnitt 6.4.3.2).

<sup>10</sup> Vgl. zur Evaluation und Validierung der entsprechenden Artefakte Abschnitt 1.2.2.

<sup>11</sup> Vgl. Heinrich/Heinzl/Roithmayr (2007): 13.

<sup>12</sup> Wissenschaftliche Kommission Wirtschaftsinformatik (2011): 1. Die Abkürzung ‚IS‘ wird hier im Original eingeführt und der Vollständigkeit halber wiedergegeben. Sie wird im weiteren Verlauf der Arbeit nicht in diesem Sinne verwendet.



Informationssysteme werden – wie in Abb. 1 dargestellt – dabei definiert als „[...] soziotechnische Systeme, die menschliche und maschinelle Komponenten (Teilsysteme) umfassen“<sup>13</sup>. Sie dienen unterschiedlichen Aufgaben im Kontext von Daten, Information und Wissen.<sup>14</sup> In ihrer Zielformulierung folgt die WI dabei einem vorwiegend gestaltungsorientierten Ansatz,<sup>15</sup> der im Rahmen der Auseinandersetzung mit den wissenschaftstheoretischen Grundlagen und Ideen der benachbarten und verwandten Disziplinen (z. B. der Betriebswirtschaft und Informatik sowie im internationalen Kontext dem *Information Systems* (IS)), für den deutschsprachigen Raum definiert wurde.<sup>16</sup>

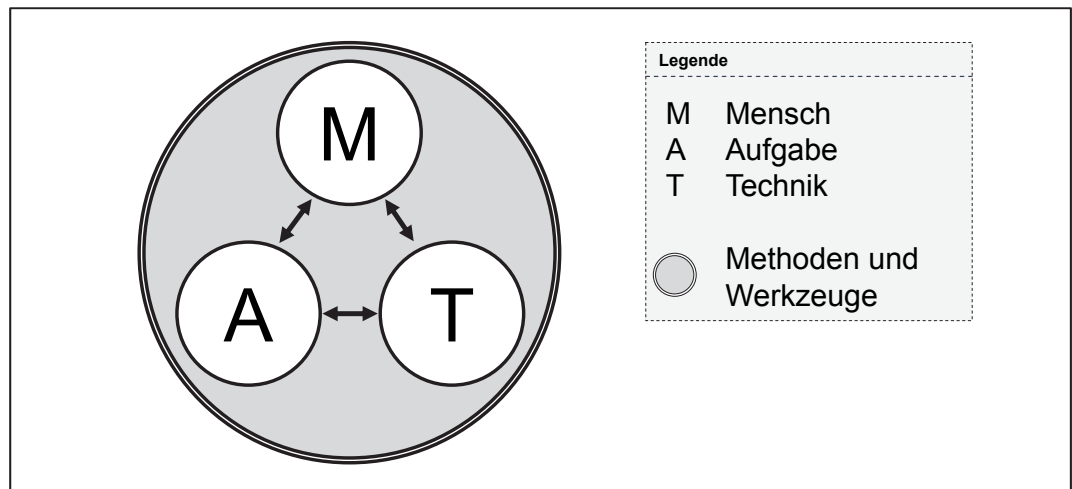


Abb. 1: Generelle Struktur eines Informationssystems

(Quelle: Heinrich/Heinzl/Roithmayr (2007): 16)

Zur Bewertung von Forschungsergebnissen der WI muss, wie generell in der Wissenschaft, Bezug genommen werden auf die Einordnung der Forschung in den Kreis möglicher Erkenntnis-, Ziel- und Methodenformulierungen, für die BECKER ET AL. einen Ordnungsrahmen vorschlagen (siehe Abb. 2).

<sup>13</sup> Wissenschaftliche Kommission Wirtschaftsinformatik (2011): 1.

<sup>14</sup> Vgl. ebd.

<sup>15</sup> Vgl. Abschnitt 1.2.1.

<sup>16</sup> Vgl. Becker et al. (2009): 5f.

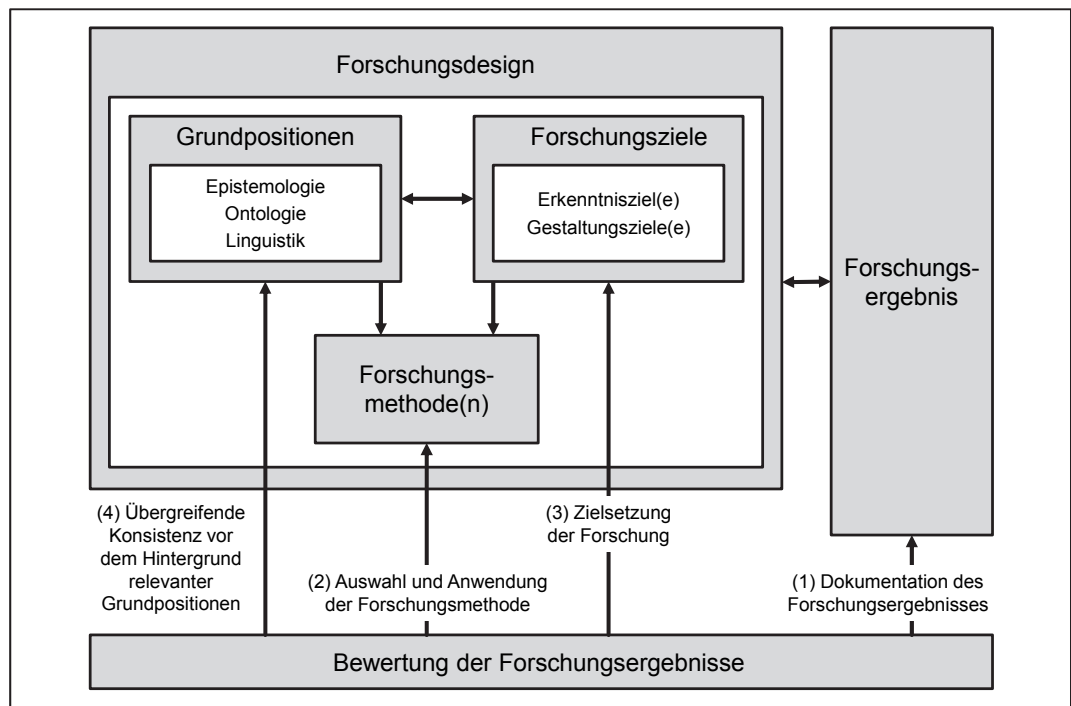


Abb. 2: Begründungsinstanzen der Bewertung von Forschungsergebnissen  
(Quelle: Becker et al. (2004): 350)

Vor diesem Hintergrund stellt die vorliegende Arbeit die Dokumentation der ihr zugrunde liegenden Forschung und des Ergebnisses dar. Die Zielformulierung erfolgte bereits in Abschnitt 1.1. Die erkenntnistheoretischen Grundlagen und Einordnungen werden im Folgenden erörtert, bevor in Abschnitt 1.2.2 die verwendeten Methoden im Kontext des Methodenspektrums der WI vorgestellt werden.

### 1.2.1 Einordnung der gestaltungsorientierten Wirtschaftsinformatik

Verstanden als Prozess, dient Wissenschaft einer systematischen Erkenntnisvermehrung,<sup>17</sup> bei der Erklärung und Gestaltung die Kernaufgaben darstellen.<sup>18</sup> Dabei lässt sich der Kanon der Wissenschaften nach verschiedenen Kriterien unterteilen. Die WI versteht sich als stark interdisziplinäre Wissenschaft, deren Fundamente in der Wirtschaftswissenschaft, dabei insbesondere in der Betriebswirtschaft, und in der Informatik liegen, womit sie zwar als wirtschaftswissenschaftliche, teilweise auch als sozialwissenschaftliche Disziplin gilt, aber stark geprägt

<sup>17</sup> Vgl. Kornmeier (2007): 4.

<sup>18</sup> Vgl. Heinrich/Heinzl/Roithmayr (2007): 18.

wird von ingenieurwissenschaftlichen Ideen und Vorgehensweisen.<sup>19</sup> Als solche wird sie als Realwissenschaft aufgefasst, die sich mit tatsächlich existierenden Phänomenen (z. B. Unternehmen, Personen und Systeme) beschäftigt.<sup>20</sup>

Zum Verständnis des erkenntnistheoretischen (epistemologischen) Hintergrunds der WI kann ein Überblick der wesentlichen Positionen beitragen (vgl. Abb. 3),<sup>21</sup> eine Einordnung ist jedoch nur vor dem Hintergrund der wesentlich umfangreicher geführten Diskussion über die gestaltungsorientierte WI im deutschsprachigen Raum möglich, die sich, trotz Anerkennung verhaltenswissenschaftlicher Methoden, damit von ihrer amerikanischen ‚Schwesterdisziplin‘ IS absetzt.<sup>22</sup>

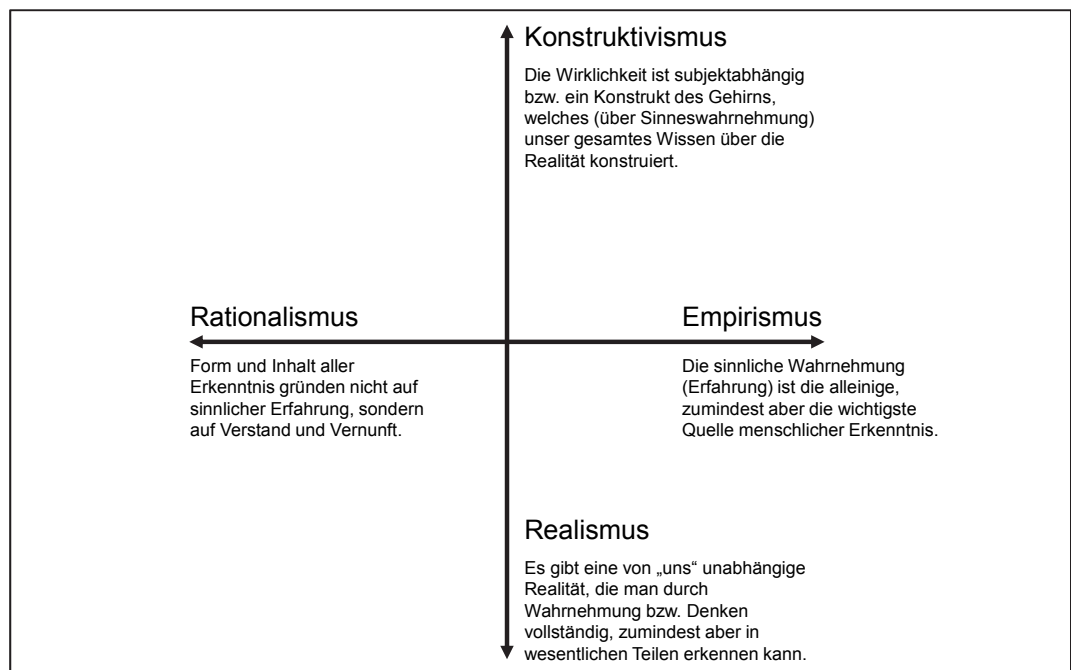


Abb. 3: Schematische Darstellung zentraler erkenntnistheoretischer Positionen (Quelle: Kornmeier (2007): 31)<sup>23</sup>

<sup>19</sup> Vgl. Heinrich/Heinzl/Roithmayr (2007): 13; Wissenschaftliche Kommission Wirtschaftsinformatik (2011): 2.

<sup>20</sup> Vgl. Kornmeier (2007): 14.

<sup>21</sup> Für eine ausführliche Gegenüberstellung der Positionen, Zwischenpositionen und historischen Entwicklung vgl. Kornmeier (2007): 31-43; sowie ergänzend Heinrich/Heinzl/Roithmayr (2007): 79-85.

<sup>22</sup> Diese Positionierung wird von der deutlichen Mehrheit der Publikationen im Umfeld der Thematik vertreten (vgl. Becker et al. (2009): 6; Herzwurm/Stelzer (2008): 2; Schauer (2011): 2; Steininger et al. (2009): 480; Wilde/Hess (2007): 280) und auch aktiv gefordert (vgl. Österle et al. (2010): 662–664).

<sup>23</sup> KORNMEIER führt die Grafik auf WILLIMCZIK zurück (vgl. Willimczik (2002): 15). Dort wird statt des Konstruktivismus der Idealismus mit bedeutungsidentischer Beschreibung verwendet.

Vor dem Hintergrund einer fortschreitenden Internationalisierung gerät die WI dabei unter Druck, ihren Fokus von der vorwiegend an der betrieblichen Praxis ausgerichteten Nützlichkeit (*Relevance*) zu verlagern oder zumindest zu erweitern und eine klar spezifizierte Methodik der Forschung selbst (*Rigor*) zu berücksichtigen.<sup>24</sup> Im Gegenzug übernimmt die IS in Teilen die Methoden einer konstruktion-orientierten Wissenschaft und führt diese unter dem Begriff *Design Science (Research)*. Die WI greift die dort entwickelten Ideen und Prinzipien wieder auf und verweist insbesondere auf HEVNER ET AL., die die Artefaktkonstruktion (Konstrukte, Modelle, Methoden und Instanziierungen) zur Lösung eines Problems als – zu dem verhaltenswissenschaftlichen Ansatz alternativen – Weg zum Verständnis eines Problems beschreiben.<sup>25</sup> RIEGE/SAAT/BUCHER formulieren dies für die WI explizit als Zusammenführung von „Erkenntnisgewinnung sowie [...] Konstruktion des Artefakts in einem Forschungsprozess“<sup>26</sup>. Die so geschaffene Verzahnung der Wissenschaften führt zu einer nicht mehr präzisen Bestimmung der verwendeten Begriffe. So werden z. B. ‚konstruktiv‘ und ‚konstruktorientiert‘ in Teilen synonym verwendet und erschweren einen Rückschluss auf die Einordnung im Rahmen der beschriebenen erkenntnistheoretischen Positionen.<sup>27,28</sup>

Eine Bestimmung des epistemologischen Hintergrunds der Forschung lässt sich so nur für spezielle Teilgebiete der WI oder Forschungsvorhaben definieren, wie es bspw. BECKER ET AL. für fachkonzeptionelle Informationsmodellierungen tun. Sie verweisen auf eine „Konsensorientierte Informationsmodellierung“<sup>29</sup>, die sie als

---

<sup>24</sup> Vgl. Becker et al. (2009): 6.

<sup>25</sup> Vgl. Hevner et al. (2004): 77. HEVNER setzt sich in seinen Arbeiten aus anglo-amerikanischer Perspektive mit den unterschiedlichen Sichtweisen auf gestaltungsorientierte Forschung auseinander und stellt z. B. ein Modell der gestaltungsorientierten Forschung vor, das den langfristigen Effekt von Artefakten berücksichtigt (vgl. Gill/Hevner (2013)). Der potenzielle Mehrwert dieser Artefakte kann dann durch eine Einordnung in das ‚DSR Knowledge Contribution Framework‘ abgeschätzt werden (vgl. Gregor/Hevner (2013): 345–347). Für die vorliegende Arbeit kann argumentiert werden, dass die Lösungsreife sehr hoch ist (Netzwerke, BI-Systeme, Analysesysteme), die Anwendungsdomäne (Kooperative BI) jedoch noch nicht sehr umfangreich erforscht ist. Der in dieser Arbeit vorgestellte Ansatz ist damit als Exaptation („exaptation“) zu verstehen, da er bekannte Lösungsansätze in eine neue Domäne transferiert.

<sup>26</sup> Vgl. Riege/Saat/Bucher (2009): 70. Die Auslassung lautet im Original „die“.

<sup>27</sup> Vgl. Maur (2009): 133, 151f.

<sup>28</sup> Auch die IS führt ähnliche Diskussionen zur korrekten Begrifflichkeit, wie sich z. B. an der Notwendigkeit zeigt, ‚scientific design‘, ‚design science‘ und ‚science of design‘ voneinander abzugrenzen (vgl. Cross (2001): 3).

<sup>29</sup> Becker et al. (2004): 351.

gemäßigt konstruktivistisch einordnen.<sup>30</sup> Auch aus Sicht von FRANK erscheint eine eindeutige Einordnung der gesamten WI nicht möglich. Er stellt für den Beleg dieser Einschätzung mehrere Ansätze (Logischer Positivismus, Kritischer Rationalismus, Kritische Theorie, Erlanger Konstruktivismus) einander gegenüber und kommt zu dem Schluss, dass keine der geprüften Schulen in der Lage ist, die Anforderungen der WI zu erfüllen, insbesondere nicht vor dem Hintergrund einer Artefaktkonstruktion als Weg zur Erkenntnisgewinnung.<sup>31</sup> Eine übergeordnete Theorie der Wirtschaftsinformatik ist damit nicht vorhanden.<sup>32</sup>

Im Sinne der wissenschaftstheoretischen Einordnung lässt sich diese Arbeit daher als gestaltungsorientiertes Werk positionieren, dessen Forschungsergebnisse durch die Auswahl und Prüfung geeigneter Methoden bewertet werden können, wie sie im folgenden Abschnitt vorgestellt werden.

### 1.2.2 Forschungsmethoden der Wirtschaftsinformatik

Vor dem Hintergrund der Diskussion um ‚Rigor vs. Relevance‘ wurden Studien durchgeführt, die für die deutsche WI offenlegen, dass die Nennung (und in noch deutlicherem Maße die Reflexion) der verwendeten Forschungsmethoden nur in einem kleinen Teil der wissenschaftlichen Publikationen stattfindet. So zeigt HEINRICH, dass in der für die deutsche WI wichtigsten Publikation<sup>33</sup> von 1999–2003 lediglich 11,0% der untersuchten 538 Forschungsbeiträge die Forschungsmethodik explizieren<sup>34</sup>. Die Fortführung der Studie für die Jahre 2004–2007 weist eine Erhöhung auf 24,7% aus, wird von den Autoren jedoch immer noch als gering im Vergleich zur angelsächsischen IS eingeschätzt.<sup>35</sup> Die Häufigkeit des Einsatzes verschiedener Methoden stellen WILDE/HESS dar, die gleichzeitig eine Einordnung nach Paradigma und Formalisierungsgrad vornehmen (vgl. Abb. 4).

---

<sup>30</sup> Während der (radikale) Konstruktivismus davon ausgeht, dass die Welt rein subjektabhängig ist, erkennt der gemäßigte Konstruktivismus eine objektive, subjektunabhängige Welt an, deren Erfassung und Beschreibung subjektabhängig ist. ‚Wahrheit‘ oder ‚Wirklichkeit‘ wird auf Basis dieser Erfassungen durch den Konsens der beobachtenden Subjekte definiert (vgl. Becker et al. (2004): 344; Navrade (2008): 6-9).

<sup>31</sup> Vgl. Frank (2006): 20f.

<sup>32</sup> Vgl. Heinrich/Heinzl/Roithmayr (2007): 84.

<sup>33</sup> Die WIRTSCHAFTSINFORMATIK weist das höchste Ranking im VHB JOURQUAL 2.1 einer WI-spezifischen, deutschsprachigen Zeitschrift auf (vgl. VHB (2011)). Sie wird seit Februar 2015 im VHB JOURQUAL 3.0 unter ihrem englischen Titel ‚Business & Information Systems Engineering (BISE)‘ geführt (vgl. VHB (2015)).

<sup>34</sup> Vgl. Heinrich (2005): 105–107.

<sup>35</sup> Vgl. Becker et al. (2009): 10, 16.

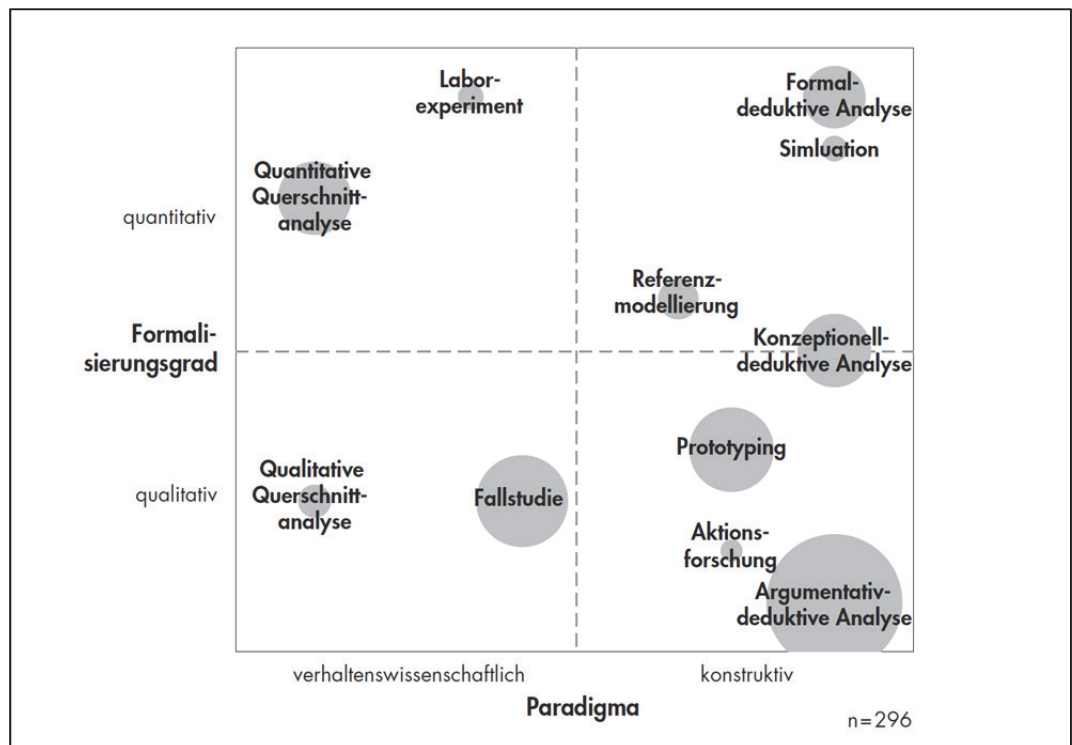


Abb. 4: Empirisch gestütztes Methodenprofil der Wirtschaftsinformatik  
(Quelle: Wilde/Hess (2007): 284)

Für die konstruktive, d. h. die gestaltungsorientierte, WI lassen sich dabei deduktive Analysen, Prototyping, Referenzmodellierung und Simulation als die am häufigsten eingesetzten Werkzeuge identifizieren. Das Verständnis dieser Methoden wird in Tab. 1 dargestellt.

Methode	Beschreibung
<b>Formal-/ konzeptionell- und argumentativ-deduktive Analyse</b>	Logisch-deduktives Schließen auf mehreren Formalisierungsstufen, z. B. mathematisch-formal, semi-formal oder rein sprachlich.
<b>Simulation</b>	Abbildung von Systemverhalten und -umwelt in Modellen. Der Erkenntnisgewinn erfolgt durch die Modellkonstruktion selbst und durch Beobachtung endogener Modellgrößen.
<b>Referenzmodellierung</b>	Induktive oder deduktive Herleitung vereinfachter und optimierter Systemabbildungen.
<b>Prototyping</b>	Entwicklung und Evaluation von Vorabversionen von Anwendungssystemen.

Tab. 1: Ausgewählte Methoden der gestaltungsorientierten Wirtschaftsinformatik  
(Quelle: Gekürzt und angepasst übernommen aus Wilde/Hess (2007): 282)

Vor dem Hintergrund der beschriebenen Verzahnung von WI und IS schlagen GERICKE/WINTER vor, im Bereich der Artefaktkonstruktion auch die Erkenntnisse

des Design Science miteinzubeziehen.<sup>36</sup> PEFFERS ET AL. stellen hierzu eine Vorgehensmethodik vor (vgl. Abb. 5). Sie beziehen sich dabei auf die *Richtlinien für Design Science in IS* von HEVNER ET AL.<sup>37</sup>

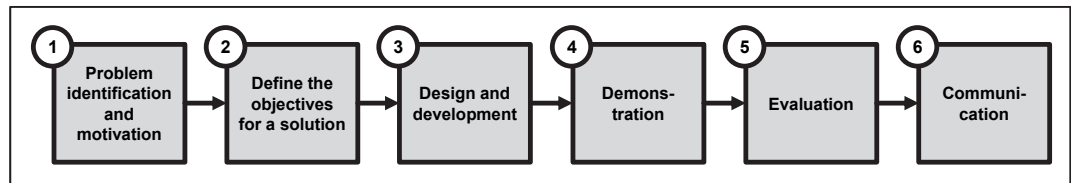


Abb. 5: Design Science Research Methodology

(Quelle: In Anlehnung an Peffers et al. (2007): 12–14)

Trotz der zunehmenden Fundierung der Methodik und des internationalen Austausches weisen aktuelle Beiträge zur Diskussion um den geeigneten Einsatz von Forschungsmethoden in der gestaltungsorientierten WI darauf hin, dass derzeit noch keine vollständige Durchdringung der Disziplin hinsichtlich einheitlicher Evaluierungs- und Rechtfertigungsmethoden stattgefunden hat, sondern eine Prüfung immer spezifisch und vorwiegend gegen Forschungslücken und weniger gegen die Realwelt vorgenommen wird.<sup>38</sup> Als sinnvoller Weg, um Forschung in der gestaltungsorientierten WI durchzuführen und zu begründen, wird die Wahl einer Methodenkombination angesehen.<sup>39,40</sup>

Die vorliegende Arbeit verwendet daher die ausgewählten, vorgestellten Methoden in unterschiedlichem Umfang. Ein argumentativ-deduktives Vorgehen wird im Rahmen der Literaturanalysen<sup>41</sup> und der Ableitung der Notwendigkeit von Gütekriterien für P2P-basierte Business-Intelligence-Netzwerke eingesetzt.<sup>42</sup> Zur Bestimmung dieser Notwendigkeit wird ein Modell verwendet, das durch seine

<sup>36</sup> Vgl. Gericke/Winter (2009): 196.

<sup>37</sup> Vgl. zu den Richtlinien Hevner et al. (2004): 82f.

<sup>38</sup> Vgl. Gericke/Winter (2009): 201–204; Riege/Saat/Bucher (2009): 82f.

<sup>39</sup> Vgl. Frank (2006): 62. Der Autor schlägt den Einsatz eines Frameworks vor, das verschiedene Methoden(-bestandteile) kombinierbar macht (vgl. Frank (2006): 46).

<sup>40</sup> Die Gemeinschaft der Wirtschaftsinformatik beschäftigt sich zudem weiterhin mit der Frage, ob der vorliegende ‚Methodenpluralismus‘ eher positiv oder negativ zu sehen ist und die Bildung eines einheitlichen Verständnisses und einer gesellschaftlichen (positiven) Wahrnehmung fördert oder behindert (vgl. Loos et al. (2013): 457–464).

<sup>41</sup> Vgl. insb. Abschnitt 3.3.1 und Abschnitte 4.2–4.4. Die Zuordnung von Literaturanalysen zur argumentativ-deduktiven Methodik erfolgt bei WILDE/HESS (vgl. Wilde/Hess (2007): 283).

<sup>42</sup> Vgl. Abschnitt 4.5.2.5 und Kapitel 5.

Abstraktheit als Referenzmodell bezeichnet werden kann.<sup>43</sup> Eine prototypische Umsetzung dient im Folgenden dazu, Veränderungen und deren Auswirkungen an einer idealtypischen und reduzierten Umsetzung des Modells durch die simulationshafte Änderung von Modellgrößen nachzuvollziehen.<sup>44</sup> Dabei werden auch hypothesengenerierende und -überprüfende Verfahren eingesetzt, um die Eignung des Modells für den gegebenen Zweck nachzuweisen. Eine Erörterung der angemessenen Verwendung der Methoden für den Forschungszweck wird in den jeweiligen Abschnitten vorgenommen.

### 1.3 Aufbau der Arbeit

Kapitel 2 stellt zunächst einen Überblick über Kooperationsmotivation und Kooperationsformen von Unternehmen dar. Der Fokus wird dabei auf unternehmensübergreifende Zusammenarbeit gelegt und es werden Spezifika bei der Zusammenarbeit in Analyseprozessen ergänzt. In Abschnitt 2.4 wird ein Fallbeispiel eingeführt, das auf real existierenden Unternehmensnetzwerken im Bereich der Automobilzulieferindustrie basiert und im Verlauf der Arbeit den betriebswirtschaftlichen Kontext der Überlegungen zu Netzwerken im Bereich der Entscheidungsunterstützung darstellt.

Die notwendigen Grundlagen für den Einsatz von Entscheidungsunterstützungssystemen in Unternehmen werden in Kapitel 3 vermittelt. Nach einer begrifflichen und fachlichen Einordnung von Business Intelligence als Leitbegriff der Diskussion wird die existierende Forschung zur Unterstützung von Gruppenentscheidungen aufgegriffen. Hierbei gilt der Terminus Collaborative Business Intelligence derzeit als maßgeblich für Kooperationsformen im Bereich BI. Für diesen wird daher eine ausführliche Literaturübersicht gegeben, die als Basis zur Klassifizierung bestehender Ansätze dient.

Kapitel 4 stellt verschiedene Probleme im Kontext von BI-Netzwerken vor und unterscheidet zu diesen mehrere Lösungsansätze. Insbesondere wird in Abschnitt 4.5 begründet, warum ein P2P-basiertes Netzwerk eine geeignete Basis für ein BI-Netzwerk schafft und welche Funktionen auf analytischer Ebene damit unterstützt werden können. Dafür wird zuvor der bisherige Stand der Forschung dieses Be-

---

<sup>43</sup> FETTKE/BROCKE weisen darauf hin, dass kein einheitliches oder eindeutiges Verständnis des Referenzmodellbegriffs existiert (vgl. Fettke/Brocke (2013)). Eine Auseinandersetzung mit dieser Thematik wird in Abschnitt 4.5.1 vorgenommen.

<sup>44</sup> Vgl. Kapitel 6.



reichs aufgearbeitet und schließlich ein einfaches Referenzmodell entwickelt, das als Grundlage für Überlegungen zu notwendigen Rahmenparametern und festlegbaren Größen P2P-basierter BI-Netzwerke dient, um die Nutzbarkeit der Netzwerke sicherzustellen.

Diese Sicherstellung ist Gegenstand des Kapitels 5. Unter Berücksichtigung graphentheoretischer Grundlagen werden dort beeinflussbare und unbeeinflussbare Parameter der Netzwerke betrachtet. Die Bewertung der konfigurierten Netzwerke im Sinne einer Nutzbarkeit für BI-Zwecke findet anhand von Gütekriterien statt, die unter Einbezug vorhandener, BI-spezifischer Kriterien entwickelt werden.

In Kapitel 6 werden eine prototypische Implementierung des Modells, ihre Architektur und unterschiedliche Modellparameterkonstellationen vorgestellt. Es erfolgen die Spezifikation der konkreten Parameterbelegungen und die Präsentation der Gütekriterienwerte als Ergebnis der durchgeführten Simulationen. Die darauf folgende Interpretation der Ergebnisse bildet die Grundlage für die Empfehlungen und Implikationen auf die Gestaltung von nutzungsfähigen P2P-basierten BI-Netzwerken.

Die Arbeit schließt in Kapitel 7 mit einem Resümee und einem Ausblick auf weitere Forschungsarbeit in diesem Bereich.

## 2 Zusammenarbeit von Unternehmen

Unternehmen können als sozio-ökonomische Systeme verstanden werden, die Güter oder Dienstleistungen erstellen und verwerten.<sup>45</sup> Die dafür notwendigen Aktivitäten fasst PORTER in einer Wertkette zusammen (vgl. Abb. 6).

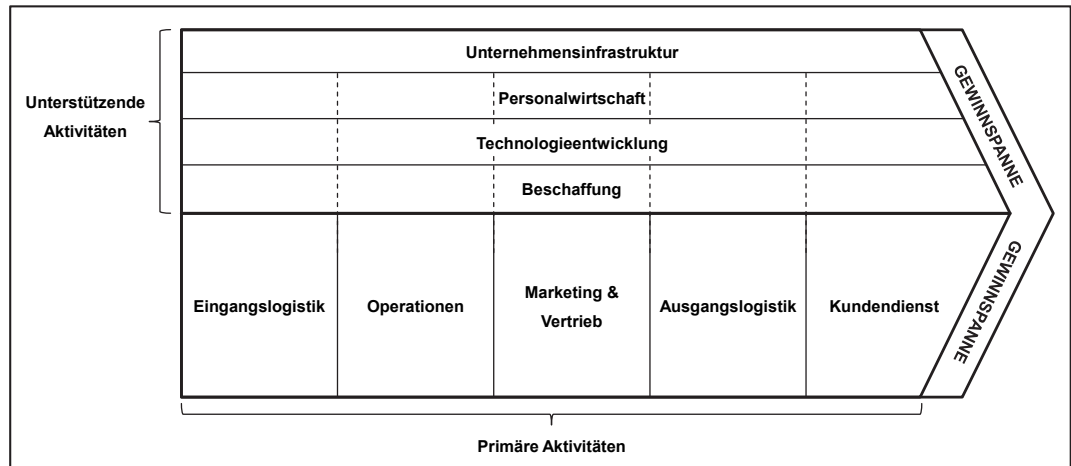


Abb. 6: Modell einer Wertkette

(Quelle: Porter (2014): 64)

Der Einsatz menschlicher Ressourcen in den einzelnen Wertaktivitäten setzt eine Zusammenarbeit der Beteiligten voraus, damit die einzelnen Aktivitäten eine zusammenhängend operierende Kette bilden können.<sup>46</sup>

Verknüpfungen bestehen aber auch zwischen den Wertketten unterschiedlicher Unternehmen.<sup>47</sup> Die daraus entstehende Wertschöpfungskette (engl.: Supply Chain) muss von den beteiligten Unternehmen aufmerksam betrachtet werden, da die Güte des in ihr abgebildeten Güter-, Werte- und Informationsflusses die Stärke der Unternehmen im Wettbewerb beeinflusst.<sup>48</sup> Die Verwaltung und Verbesserung dieser Ketten und der Kooperation in ihnen wird in der Literatur umfangreich diskutiert und als *Supply Chain Management* bezeichnet.<sup>49</sup>

Im Folgenden werdem der Kooperationsbegriff abgegrenzt und auf Basis unterschiedlicher betriebswirtschaftlicher Theorien die Motivation für Kooperationen

<sup>45</sup> Vgl. Hutzschenreuter (2009): 7.

<sup>46</sup> Vgl. Porter (2014): 67, 78–80.

<sup>47</sup> Diese werden als ‚vertikale Verknüpfungen‘ bezeichnet (vgl. Porter (2014): 81f.).

<sup>48</sup> Kurbel (2013): 222.

<sup>49</sup> Vgl. Aliche (2005): 3–5; Kuhn/Hellingrath (2002): 10; Kurbel (2013): 223; sowie die Ausführungen in Abschnitt 2.3.2.

herausgearbeitet. Im Anschluss werden Merkmale und zentrale Konzepte von Kooperationen identifiziert.

Im operativen Geschäft ist die Notwendigkeit zur Kooperation bereits anhand der Wertkette direkt nachvollziehbar, da der (Waren-)Ausgang des einen Unternehmens unmittelbar den (Waren-)Eingang anderer Unternehmen beeinflusst. Es besteht aber auch in den zunächst aus Unternehmenssicht autonom erscheinenden Führungsfunktionen die Möglichkeit, unternehmensübergreifend zusammenzuarbeiten.<sup>50</sup> Abschnitt 2.4 liefert ein Fallbeispiel einer Kooperation in Planungs-, Entscheidungs- und Analyseprozessen, in denen Entscheidungsunterstützung besonders relevant wird.

## 2.1 Definition und Motivation

Der Begriff der Kooperation folgt in der (betriebswirtschaftlichen) Literatur keiner eindeutigen Definition. Eines der kürzesten und allgemeingültigsten Verständnisse wird einleitend von FONTANARI genannt: „Kooperation bedeutet Zusammenarbeit.“<sup>51</sup> Die Mehrzahl der Autoren von Werken, die sich zentral mit Kooperation beschäftigen, weist zunächst auf eine große Menge synonym gebrauchter Ausdrücke hin.<sup>52</sup> Auch in der angelsächsischen Literatur findet sich keine einheitliche begriffliche Verwendung oder Beschreibung.<sup>53</sup> RUPPRECHT-DÄULLARY weist darauf hin, dass dort ‚cooperation‘ und ‚co-operation‘ vorwiegend in volkswirtschaftlichem Zusammenhang verwendet werden.<sup>54</sup> Es wird zu meist auf die notwendigen Eigenschaften einer Zusammenarbeitsbeziehung abgestellt, die eine Kooperation im betriebswirtschaftlichen Sinne ausmachen.<sup>55</sup>

---

<sup>50</sup> Als Führungsfunktionen werden hier Planung, Steuerung und Kontrolle verstanden (vgl. Hahn/Hungenberg (2001): 48), wobei die dafür notwendigen Analyse- und Entscheidungsprozesse mit einbezogen werden (vgl. Hammer (2011): 35–39).

<sup>51</sup> Fontanari (1996): 34.

<sup>52</sup> Vgl. z. B. Schwerk (2000): 26–31. Eine Literaturübersicht verschiedener Zusammenstellungen findet sich bei RUPPRECHT-DÄULLARY (vgl. Rupprecht-Däullary (1994): 5). BALLING stellt eine Liste zusammen, in der sich Begriffe unterschiedlicher Abstraktionsniveaus finden, darunter ‚Koalition‘ – ein Begriff, den auch PORTER nutzt (vgl. Porter (2014): 89) –, ‚Joint Venture‘ und ‚Quasi-Integration‘ (vgl. Balling (1998): 13). Vgl. hierzu auch die Hinweise zu dem Verhältnis von ‚Kollaboration‘ und ‚Kooperation‘ im Rahmen dieser Arbeit in Fußnote 6.

<sup>53</sup> Vgl. Schwerk (2000): 26; Zentes/Swoboda/Morschett (2003): 5.

<sup>54</sup> Vgl. Rupprecht-Däullary (1994): 11.

<sup>55</sup> Vgl. Balling (1998): 15; Fontanari (1996): 35; Rupprecht-Däullary (1994): 16; Schwerk (2000): 28; Sell (2002): 3.

Diese Arbeit folgt zur klaren Abgrenzung der präziseren Definition von SELL und definiert Kooperation als

- „- eine zwischenbetriebliche Zusammenarbeit
- zwischen mindestens zwei rechtlich und wirtschaftlich in den nicht von der Kooperation betroffenen Gebieten selbständigen Unternehmen
- zur gemeinsamen Durchführung von Aufgaben,
- die in der Regel auf mittlere bis längere Frist angelegt ist.“<sup>56,57</sup>

Das Eingehen einer solchen Kooperation führt zu einer Bindung und erfordert eine Motivation – insbesondere dann, wenn der Kooperationspartner zumindest teilweise im Wettbewerb zu dem eigenen Unternehmen steht. Die Bereitschaft zur Kooperation kann dabei über eine Vielzahl betriebswirtschaftlicher Theorien erklärt werden.<sup>58</sup>

Unter der Annahme, dass insbesondere, wenn führungsrelevante Informationen Bestandteil des Austausches einer Kooperation werden, ohne diese keine vollständige Markttransparenz vorliegt, d. h. auch kein vollkommener Markt existiert, kann z. B. die Transaktionskostentheorie herangezogen werden. Transaktionskosten entstehen für die mit der Durchführung verbundenen Aktivitäten, wenn Handlungs- und Verfügungsrechte von Gütern übertragen werden.<sup>59</sup> In der Transaktionskostentheorie steht die Betrachtung dieser Transaktionen im Mittelpunkt, wobei ein Informationsungleichgewicht und ein Mangel an Vorhersehbarkeit aller in Zukunft entstehender Kosten die Teilnehmer einer (potenziellen) Transaktion ggf. in die Lage versetzt, Vorteile zu Lasten anderer Teilnehmer zu realisieren.<sup>60</sup> Unterschiedliche Kooperationsformen von Unternehmen lassen sich dann auf verschiedene erwartete Transaktionskostenkonstellationen zurückführen. So kann z. B. im Falle mittelhoher Transaktionskosten eine Allianz verschiedener Unternehmen gerechtfertigt werden, weil diese die (ungewünschten) Transaktionskos-

<sup>56</sup> Sell (2002): 3.

<sup>57</sup> Hier sei die Einschränkung der vierten Eigenschaft hervorgehoben, die eine mehr als temporäre Zusammenarbeit nur ‚in der Regel‘ vorsieht. ZENTES/SWOBODA/MORSCHETT verweisen in ihrer Definitionsfindung, in der sie ebenfalls auf SELL Bezug nehmen, noch stärker explizit darauf, dass die Dauer der Zusammenarbeit keine entscheidende Rolle bei der Begriffsbestimmung von ‚Kooperation‘ einnimmt (vgl. Zentes/Swoboda/Morschett (2005b): 5).

<sup>58</sup> Vgl. Swoboda (2005): 56f. Die dortige Übersicht wird durch diverse Beiträge in dem gleichen Sammelwerk ergänzt und detailliert (vgl. Zentes/Swoboda/Morschett (2005a)). Eine nicht theoriegebundene Klassifikation nimmt z. B. STRAUBE vor, der volks- und betriebswirtschaftliche sowie gesellschafts- und mittelstandspolitische Motive unterscheidet (vgl. Straube (1972): 7–17).

<sup>59</sup> Vgl. Picot (1991): 145.

<sup>60</sup> Vgl. Picot (1991): 147.

ten senken kann, dabei aber noch keine vertikale Integration erfordert, die wiederum mit einer erheblichen (ungewünschten) Veränderung der Unternehmensstruktur einherginge.<sup>61</sup> Weitere Konstellationen, z. B. für Joint Ventures,<sup>62</sup> finden sich u. a. bei RAMANATHAN/SETH/THOMAS.<sup>63</sup>

Als zweite, von ihrem Ursprung her mathematische Erklärungsidee wird im Folgenden auf die Spieltheorie eingegangen. Diese wird im Rahmen dieser Arbeit auch bei der Einordnung der Business Intelligence in die allgemeine Entscheidungsunterstützung verwendet,<sup>64</sup> und ist mit ihrem Fokus auf Entscheidungssituationen besonders geeignet, die grundsätzliche Idee einer Kooperation zu beschreiben.<sup>65</sup> MAGIN/HEIL/FÜRST definieren die Spieltheorie als „ein Konzept zur Optimierung von Entscheidungen, deren Ergebnis von der Reaktion der Konkurrenten beeinflusst wird.“<sup>66</sup> Als einfache Erklärung der Vor- und Nachteile einer (fehlenden) Kooperation mehrerer kooperationsfähiger potenzieller Partner dient zumeist das ‚Gefangenendilemma‘. Es besteht in diesem eine Situation, in der zwei Partner die unabhängige Wahl haben, zu kooperieren oder nicht zu kooperieren (z. B. durch Erfüllung eines Vertrages, Leistung einer Zahlung, etc.). Beide können durch Kooperation einen Gewinn<sup>67</sup> erzielen, der für jedes Unternehmen nur dann erhöht werden kann, wenn es nicht kooperiert, der Partner hingegen schon. Es gilt folgender Zusammenhang für den Gewinn eines Unternehmens.<sup>68</sup>

$$G^{NK,K} > G^{K,K} > G^{NK,NK} > G^{K,NK}$$

mit $G^{NK,K}$	Gewinn bei Nicht-Kooperation und Kooperation des Partners
$G^{K,K}$	Gewinn bei beidseitiger Kooperation
$G^{NK,NK}$	Gewinn bei beidseitiger Nicht-Kooperation
$G^{K,NK}$	Gewinn bei Kooperation und Nicht-Kooperation des Partners

<sup>61</sup> Vgl. Gulati (1995): 87.

<sup>62</sup> Vgl. hierzu die Beschreibungen unterschiedlicher Kooperationsformen in Abschnitt 2.3.

<sup>63</sup> Vgl. Ramanathan/Seth/Thomas (1997): 57, 65f. Weitere Beispiele sowie einen Vergleich der Transaktionskostentheorie mit der Ressourcentheorie unter besonderer Berücksichtigung strategischer Netzwerke liefern auch DAS/TENG (vgl. Das/Teng (2000): 31–36).

<sup>64</sup> Vgl. Kapitel 3.

<sup>65</sup> Die Spieltheorie wird zurückgeführt auf NEUMANN (vgl. Neumann/Morgenstern (2004)). Für eine Einführung vgl. z. B. Wölfe (2014): 179–184; sowie Magin/Heil/Fürst (2005): 123–130.

<sup>66</sup> Magin/Heil/Fürst (2005): 124.

<sup>67</sup> Gemeint ist hier kein Gewinn in betriebswirtschaftlicher Definition, sondern lediglich eine nicht näher spezifizierte erstrebenswerte Verbesserung des eigenen Zustands.

<sup>68</sup> Vgl. Bühler/Jäger (2005): 100. Die Symbole wurden für die vorliegende Arbeit verändert.

Da beide Partner den Gewinn maximieren wollen, liegt zunächst eine Nicht-Kooperation nahe. Die „Intelligenz“<sup>69</sup> der Beteiligten lässt das entsprechende Vorgehen des anderen allerdings antizipieren, sodass eine Kooperation vorteilhaft erscheinen kann. Die nicht-formale Absprache dieses Einhaltens der gegenseitigen Kooperation wird als *Kollusion* bezeichnet und im Bereich der (teilweise illegalen) Kartellbildung<sup>70</sup> mit besonderem Interesse verfolgt.<sup>71</sup>

Während sich aus den Theorien die grundsätzliche Bereitschaft zur Kooperation motivieren lässt, wird in der Literatur auch auf Basis von Erfahrungen aus der Praxis die schlichte Erkenntnis formuliert, dass eine Nutzung von Allianzen und entsprechenden Möglichkeiten der Zusammenarbeit in verschiedenen Formen notwendig geworden ist.<sup>72</sup>

## 2.2 Kriterien für Unternehmenskooperationen

Im vorherigen Abschnitt wurden im Kontext der Kooperationsdefinition einzelne Aspekte von Kooperationen herausgearbeitet. Die Literatur unterscheidet zur Erklärung anhand unterschiedlicher Merkmale, Dimensionen oder Kriterien<sup>73</sup> eine Vielzahl differenzierter Ausprägungen. Diese werden im Folgenden dargestellt. Häufig genannte und für unterschiedliche Kooperationsformen prägende Kriterien werden dabei im Anschluss nochmals separat behandelt.

---

<sup>69</sup> Magin/Heil/Fürst (2005): 124. Intelligenz wird hier als feststehender Begriff im Rahmen der Definition der Spieltheorie als Eigenschaft der Beteiligten so formuliert wie im Textzusammenhang verwendet.

<sup>70</sup> Insbesondere problematisch sind Preiskartelle, die die Findung eines ‚fairen‘ Preises für Anbieter und Nachfrager zu Ungunsten der Nachfrager behindern: „Das Ergebnis einer expliziten Preisabsprache zwischen den Unternehmen wird als Preiskartell bezeichnet. In einer Preisabsprache werden sich die Beteiligten auf den Monopolpreis einigen, der den gemeinsamen Gewinn maximiert.“ (Pfähler/Wiese (2008): 87).

<sup>71</sup> Vgl. Bühler/Jäger (2005): 99.

<sup>72</sup> Vgl. z. B. Berg/Schmitt (2003): 282f.; Das/Teng (2000): 32; Gerlach/Cleophas/Kliwer (2013): 147; für Supply Chains z. B. Arshinder/Kanda/Deshmukh (2011): 40, 45f.; sowie die jeweils dort angegebenen Literaturhinweise.

<sup>73</sup> Während ein *Merkmal* ein „charakteristisches, unterscheidendes Zeichen, an dem eine bestimmte Person, Gruppe oder Sache, auch ein Zustand erkennbar wird“ (Duden (2015d)) beschreibt, ist ein *Kriterium* ein „unterscheidendes Merkmal als Bedingung für einen Sachverhalt, ein Urteil, eine Entscheidung“ (Duden (2015e)) und wird damit im Kontext von Entscheidungsfindungen behandelt. Die Literatur behandelt die Begriffe nur teilweise entsprechend getrennt und vielmehr synonym, so z. B. in Balling (1998): 152, 162. Um den verschiedenen Quellen gerecht zu werden und in Ermangelung einer Notwendigkeit, für das Verständnis von Kooperationen die Begriffe zu trennen, wird hier ebenfalls eine Gleichwertigkeit der Ausdrücke angenommen. Im Weiteren wird von Kriterien gesprochen, da eine Wahlmöglichkeit bei der Ausgestaltung von Kooperationen angenommen wird.

### 2.2.1 Ansätze in der Literatur

Zur Ermittlung der vorherrschenden Merkmale zur Klassifikation von Kooperationsarten werden unterschiedliche Ansätze betrachtet, die explizit Kriterien aufzeigen und diese zumeist in Form eines morphologischen Kastens darlegen. SCHOLTA weist in diesem Zusammenhang darauf hin, dass die morphologische Methode, d. h. die Zuordnung von möglichen Ausprägungen zu den Einzelbestandteilen einer Problemstellung, zwar eine „relativ vollständige und systematische Beschreibung des Gesamtproblems“<sup>74</sup>, jedoch nicht „eine abschließende und vollständige Abbildung des Problems ‚Kooperation‘“<sup>75</sup> bieten kann. Aus sechs Publikationen wurden die entsprechenden Vorarbeiten zusammengeführt<sup>76</sup> und so mehrfach genannte Kriterien (ggf. mit unterschiedlichen Ausprägungen) in einem morphologischen Kasten synthetisiert.<sup>77</sup> Abb. 7 zeigt die neun Kriterien, die in mindestens drei der Publikationen genannt wurden. Die zugehörige Zuordnung liefert Tab. 2.<sup>78</sup> Die Kriterien mit weniger Nennungen sowie deren Ausprägungen werden in Tab. 3 ergänzt. Das Ziel dieses Vorgehens ist die Berücksichtigung zentraler Kriterien im Kontext von Entscheidungsunterstützung.<sup>79</sup>

---

<sup>74</sup> Scholta (2005): 22.

<sup>75</sup> Ebd.

<sup>76</sup> Dabei handelt es sich um: Balling (1998): 162; Fontanari (1996): 40; Hirschmann (1998): 27; Rupprecht-Däullary (1994): 19; Schwert (2000): 32; Zentes/Swoboda/Morschett (2005b): 22.

<sup>77</sup> Das Literaturangebot allein im deutschsprachigen Raum zu Kooperationen und entsprechenden Kriterien ist erheblich größer. Da die Kooperation als solche nicht den Kern der vorliegenden Arbeit bildet, wurde auf eine ‚erschöpfende‘ Literaturrecherche verzichtet (zu einem möglichen Vorgehen einer solchen und der Abgrenzung unterschiedlich ausführlicher Arten vgl. Abschnitt 3.3.1.1). Die gefundenen Kriterien sowie die Identifikation der wichtigsten Kriterien werden aber durch weitere Literatur gestützt (vgl. z. B. Killich (2011): 18–20; Liebscher (2012): 166), die sich auch auf die hier zugrunde gelegten Arbeiten bezieht (vgl. Liebscher (2012): 166; Osiecka (2006): 9f.). Eine umfangreiche Aufstellung, die darauf hinweist, dass die Kriterien bereits seit mehreren Jahrzehnten in dieser oder ähnlicher Form verwendet werden, liefert WURCHE mit einer Literaturbetrachtung der (deutschsprachigen) Forschung zum Thema (vgl. Wurche (1994): 133).

<sup>78</sup> Zuordnungen werden durch eine Graufärbung der entsprechenden Zellen markiert.

<sup>79</sup> Vgl. hierzu Kapitel 3, insbesondere die Klassifikation von unterschiedlichen Ansätzen der Collaborative Business Intelligence in Abschnitt 3.3.3.

Kriterien	Ausprägungen			
Richtung / Wirtschaftsstufe	Horizontal	Vertikal	Diagonal / Lateral / Konglomerat	
Bindung / Vertragsform	Formal	Informell	Stillschweigend	Gewohnheitsrechtlich
	Kodifiziert		Nicht kodifiziert	
Form / Beteiligungsverhältnis	Lizenzvereinbarung	Managementvertrag	Franchising	Joint Venture
	Minderheit	Mehrheit	Parität	Keine Beteiligung
Fristigkeit / Dauer / Stabilität	Einmalig	Sporadisch	Regelmäßig	Dauerhaft
	Projektbezogen		Projektübergreifend	
	Kurzfristig	Mittelfristig	Langfristig	Unbefristet
Intensität / Art der Verflechtung	Absprachenkooperation		Austauschkooperation	Gemeinschaftskooperation
	Austausch von Informationen und Ereignissen	Koordiniert arbeitsteiliges Vorgehen	Gemeinschaftliches Vorgehen	Gemeinschaftlich getragene Organisation
Unternehmensbereiche / Funktionen	Primäraktivitäten		Sekundäraktivitäten	
	F&E	Beschaffung	Produktion	Weitere Funktionen
Anzahl der Partner / Koordinationsstruktur	Bilateral	Trilateral	Einfache Netzwerke	Komplexe Netzwerke
	Niedrig		Hoch	
Raumaspekt / Partnerherkunft	Lokal	Regional	National	International
Zielsysteme / Zielkompatibilität	Identisch	Komplementär		Neutral
	Hoch		Niedrig	

Abb. 7: Synthese morphologischer Kästen und Kriteriumsbeschreibungen für unternehmensübergreifende Kooperation  
(Quelle: Eigene Darstellung)

Kriterium	Zentes/Swoboda/ Morschett (2005)	Schwerk (2000)	Balling (1998)	Hirschmann (1998)	Fontanari (1996)	Rupprecht- Däullary (1994)	Anzahl Nennungen
Richtung / Wirtschaftsstufe							6
Bindung / Vertragsform							5
Form / Beteiligungsverhältnis							5
Fristigkeit / Dauer / Stabilität							5
Intensität / Art der Verflechtung							5
Unternehmensbereiche / Funktionen							5
Anzahl der Partner / Koordinationsstruktur							4
Raumaspekt / Partnerherkunft							4
Zielsysteme / Zielkompatibilität							3

Tab. 2: Behandlung von Kriterien zur Unterscheidung von Kooperationen in unterschiedlichen Publikationen  
(Quelle: Eigene Darstellung)



Kriterium	Ausprägungen
<b>Organisationsintensität / Dichte</b>	Repetitiv, Initiativ, Komplex / Situationsbedingt
<b>Ressourcenprofil der Partner</b>	X-Allianz, Y-Allianz, Mehrfunktionen
<b>Vertrauen zwischen den Partnern</b>	Hoch, Niedrig
<b>Zentralität / Determiniertheit</b>	Starke Hierarchie, Schwache Hierarchie
<b>Allgemeine Kooperationserfahrung</b>	Hoch, Niedrig
<b>Branchenzugehörigkeit</b>	[Keine Ausprägungen genannt]
<b>Diversität</b>	[Keine Ausprägungen genannt]
<b>Eigenschaften der Partner</b>	[Keine Ausprägungen genannt]
<b>Erfolg einer Kooperation</b>	Hoch, Niedrig
<b>Führungspartizipation</b>	Vorhanden, Nicht Vorhanden
<b>Integration der DV-Systeme</b>	Hoch, Niedrig
<b>Interdependenz</b>	Hohe Abhängigkeit, Niedrige Abhängigkeit
<b>Konnektivität</b>	Orientiert an direkten/indirekten Beziehungen
<b>Kulturelle und organisationsspezifische Kompatibilität der Partner</b>	Hoch, Niedrig
<b>Offenheit</b>	Hohe / Niedrige Ein-/Austrittsbarrieren
<b>Organisiertheit</b>	Hoch, Niedrig
<b>Parentalbeziehung</b>	Zulieferer, Kunde, Mitbewerber, Neutral, Staat / NGO
<b>Rechtsstatus der Kooperation</b>	Eigenständig, Abhängig
<b>Rechtsstatus der Parentalorganisation</b>	Eigenständig, Abhängig
<b>Redundanz</b>	[Keine Ausprägungen genannt]
<b>Ressourceneinbringung</b>	Tangible Assets, Intangible Assets
<b>Risiko- bzw. Unsicherheitsgrad</b>	Hoch, Niedrig
<b>Risikoverteilung</b>	Aliquot auf alle, Auf einige, Paritätisch
<b>Sichtbarkeit</b>	Offen, Verdeckt
<b>Verteilung d. Kooperationserfolgs</b>	[Keine Ausprägungen genannt]
<b>Zielausrichtung</b>	Operativ, Taktisch, Strategisch

Tab. 3: Weitere Kriterien zur Unterscheidung von Kooperationsarten (weniger als drei Nennungen)

(Quelle: Eigene Darstellung)

Eine zentrale Fragestellung bei der Klassifizierung von Kooperationen ergibt sich demnach durch die Kooperationsrichtung. Bei den weiteren Kriterien lassen sich in Teilen (starke) Interdependenzen erkennen. So ist bspw. bei einer Kapitalbeteiligung („Form“) das Vorliegen einer formalen, kodifizierten Bindung in hohem Maße wahrscheinlich. Vor dem thematischen Hintergrund dieser Arbeit werden zur Einordnung des Fallbeispiels<sup>80</sup> insbesondere die Kooperationsrichtung, die Kooperationsform sowie die Kooperationsintensität im Folgenden weiter erläutert. Die weiteren Kriterien sind zwar grundsätzlich im Rahmen der Entscheidungsunterstützung in Netzwerken zu behandeln, können aber durch die Fragestellung

<sup>80</sup> Vgl. Abschnitt 2.4.

zumeist eindeutig benannt werden. So ist die Anzahl der Partner bei der Betrachtung von Netzwerken bereits im Vorhinein als ‚hoch‘, bzw. ‚Netzwerk‘ definiert. Die betroffenen Unternehmensbereiche werden durch den Fokus auf Entscheidungsunterstützung auf wertkettenübergreifende Aktivitäten festgelegt.<sup>81</sup> Der Raumaspekt spielt zunächst keine tragende Rolle und die Fristigkeit wird durch die Betrachtung möglichst autonomer, ggf. nur kurzfristig interagierender Partner entsprechend der konkreten Umsetzung überlassen.

### 2.2.2 Kooperationsrichtung

Die meistvorgenommene Unterteilung von Arten der Kooperation findet sich bei der Betrachtung des Konzeptes der Wertschöpfungskette, d. h. dem Gesamtprozess der Erzeugung eines Gutes aus Rohstoffen bis zum Verbrauch des Kunden.<sup>82</sup> Betrachtet wird ein Unternehmen als ein Glied dieser Kette.

Eine *vertikale* Kooperation findet dann zwischen dem betrachteten Unternehmen und einem oder mehreren direkt vor- oder nachgelagerten Unternehmen statt.<sup>83</sup> Ziel ist es, die Leistungstiefe des eigenen Unternehmens zu verringern, um einen Fokus auf die Funktionen des eigenen Unternehmens zu legen, die gut beherrscht werden (‚Kernkompetenzen‘).<sup>84</sup> Zwischen den beteiligten Unternehmen besteht daher keine direkte Konkurrenz. Im Gegenzug ist durch die Bindung der Partner und den Verzicht auf die Erbringung von Vor- oder Nachleistungen eine Abhängigkeit voneinander gegeben.<sup>85</sup> Das Konzept der ‚Just-In-Time‘(JIT)-Lieferung sieht bspw. eine Belieferung des in der Kette nachgelagerten Unternehmens vor, die möglichst synchron zu dessen Produktion, d. h. der Verarbeitung des angelieferten Rohstoffs (oder allgemein: Input-Guts) stattfindet, was die Abhängigkeit

---

<sup>81</sup> Die Einschränkung auf eine lineare Struktur einer schriftlichen Ausarbeitung der Thematik bedingt, dass zwei wechselseitig aufeinander Bezug nehmende Kapitel, wie hier die Ausführungen zu Kooperationen (Kapitel 2) und zu Entscheidungsunterstützung (Kapitel 3), teilweise eine Vorwegnahme zukünftiger Erklärungen erfordern. An dieser Stelle sei daher für weitere Ausführungen auf Kapitel 3 verwiesen und im Weiteren nur auf die bereits beschriebene Motivation der Grundthematik (vgl. Abschnitt 1.1) Bezug genommen.

<sup>82</sup> Neben dem Güterfluss werden üblicherweise auch der Werte-/Geldfluss und der Informationsfluss betrachtet (vgl. die einleitenden Worte zu Kapitel 2). Das Konzept beschreibt zudem in der Regel einen ‚idealisierten‘ Wertschöpfungsprozess auf Basis des produzierenden Gewerbes (vgl. Kuhn/Hellingrath (2002): 10). KURBEL weist darauf hin, dass das Konzept der Wertschöpfungskette der netzwerkartigen Verflechtung vieler Unternehmen nicht gerecht wird (vgl. Kurbel (2013): 222; sowie die Ausführungen in Abschnitt 2.3.2).

<sup>83</sup> Denkbar sind auch mehrere Stufen, diese müssen dann aber durchgängig miteinander verbunden sein (vgl. Morschett (2005): 393).

<sup>84</sup> Vgl. Morschett (2005): 393.

<sup>85</sup> Vgl. Schwerk (2000): 35.

von (pünktlichen) Lieferungen verdeutlicht.<sup>86</sup> Eine rückwärts gelagerte Kooperation aus Sicht des betrachteten Unternehmens bezieht sich dementsprechend auf die Integration vorhergehender Betriebe und wird im Zusammenhang von ‚make-or-buy‘-Entscheidungen diskutiert.<sup>87</sup> Eine vorwärts gelagerte Integration kann z. B. dazu dienen, Zwischenhändler nachfolgender Stufen aus dem Prozess auszuschließen, um die Gewinnmargen der produzierten Güter bis hin zum Endkunden vollständig selbst zu realisieren.<sup>88</sup>

Eine *horizontale* Kooperation wird zwischen Unternehmen der gleichen Wertschöpfungsstufe geschlossen. In der Literatur wird dabei üblicherweise angenommen, dass die Unternehmen in der gleichen oder in einer artverwandten Branche tätig sind, eine Konkurrenz muss aber nicht notwendigerweise gegeben sein.<sup>89</sup> Im Gegensatz zur vertikalen Kooperation wird die Leistungstiefe nicht verändert. Ziel ist es, vorhandene Stärken und Schwächen der Beteiligten durch die (gemeinsame) Übernahme von Tätigkeiten sinnvoll auszugleichen. Insbesondere für kleine und mittlere Unternehmen (KMU)<sup>90</sup> besteht dabei die Möglichkeit, durch Synergieeffekte Kosteneinsparungen zu realisieren.<sup>91</sup> BALLING weist darauf hin, dass im Kontext von Kartellabwägungen bei der Unternehmensübernahme oder -fusion zur Realisierung einer horizontalen Kooperation zwischen solchen ‚mit‘ und ‚ohne Produktausweitung‘, d. h. der Verbreiterung des Sortiments hergestellter Güter, unterschieden wird.<sup>92</sup>

Sofern die beteiligten Unternehmen nicht direkte Vor-/Nachgänger einer Wertschöpfungskette sind und zusätzlich erkennbare Differenzen in der Branchenzugehörigkeit aufweisen, wird von einer *diagonalen*<sup>93</sup> (alternativ: *lateralen*<sup>94</sup> oder

---

<sup>86</sup> Vgl. Fontanari (1996): 55f.

<sup>87</sup> Vgl. Irle (2011): 1.

<sup>88</sup> Vgl. Balling (1998): 42.

<sup>89</sup> Vgl. z. B. Morschett (2005): 392; Paprottka (1996): 120.

<sup>90</sup> Zur Einordnung eines Unternehmens in die Klasse der KMU werden im Rahmen dieser Arbeit die Empfehlungen der Kommission der Europäischen Gemeinschaften vom 06.05.2003 angewendet. Ein Unternehmen gilt daher als KMU, wenn es weniger als 250 Mitarbeiter beschäftigt und (a) einen Jahresumsatz von nicht mehr als 50 Mio. EUR erzielt oder (b) die Jahresbilanzsumme 43 Mio. EUR nicht übersteigt (vgl. Kommission der Europäischen Gemeinschaften (2003): Anhang, Titel I, Artikel 2).

<sup>91</sup> Vgl. Paprottka (1996): 120.

<sup>92</sup> Vgl. Balling (1998): 42.

<sup>93</sup> Vgl. Schwerk (2000): 35.

<sup>94</sup> Vgl. Rupprecht-Däullary (1994): 20; Morschett (2005): 393.

konglomeraten<sup>95</sup>) Kooperation gesprochen. Hier wird eine Zusammenarbeit vor allem im Bereich unterstützender Bereiche vorgesehen, in denen sich Synergiepotenziale heben lassen, sofern ähnliche Standards bereits erreicht wurden.<sup>96</sup>

Für eine erste Einschätzung der Verteilung der Kooperationsarten kann eine beispielhafte empirische Untersuchung der Automobilindustrie herangezogen werden. Diese zeigt, dass 60% der Befragten, die Kooperationen eingehen, vertikale Kooperationen eingehen, die verbleibenden 40% horizontale Kooperationen.<sup>97</sup> Diagonale Kooperationen wurden nicht abgefragt, wären aber auf Grund der Untersuchung nur einer Branche auch nicht zweckdienlich.<sup>98</sup>

### 2.2.3 Kooperationsform

Bei der Betrachtung der Kooperationsform wird in der Regel von einer vertraglich festgelegten, d. h. kodifizierten, Basis im Sinne des Kriteriums ‚Bindung‘ ausgegangen.<sup>99</sup> Dies liegt nahe, da nicht-vertragliche Kooperationen in wissenschaftlichen Untersuchungen schwer zu handhaben sind. Ihre Struktur wird zwangsläufig bei rechtlich nicht zulässigen Kooperationen (‚Kartellen‘) gewählt, um eine Überprüfung zu erschweren. Entsprechende Untersuchungsmethoden wären daher nur schwer durchzusetzen.<sup>100</sup>

Eine Grundsatzentscheidung bei der Wahl der Kooperationsform ist vor diesem Hintergrund diejenige, ob eine Kapitalverflechtung der beteiligten Unternehmen stattfinden soll oder nicht – und falls ja, in welchem Maße und Verhältnis. Eine Beteiligung widerspricht dabei nicht dem Gedanken der Selbständigkeit, da diese zum einen nicht notwendigerweise rechtlich verändert wird, zum anderen nur in den nicht von der Kooperation betroffenen Bereichen gefordert wird<sup>101</sup> und zum Dritten sogar als stabilisierendes Instrument zur Kooperationsbildung betrachtet werden kann.<sup>102</sup> Sie ist damit von der vollständigen Übernahme von Unternehmen

---

<sup>95</sup> Vgl. Fontanari (1996): 65.; Schwert (2000): 35.

<sup>96</sup> Vgl. Paprotka (1996): 125.

<sup>97</sup> Vgl. Lange (2011): 14f.

<sup>98</sup> Weitere Ausführungen in Bezug auf die Automobilindustrie und die dort gelebten oder denkbaren Kooperationen finden sich bei der Beschreibung des Fallbeispiels in Abschnitt 2.4.

<sup>99</sup> Vgl. Abschnitt 2.2.1.

<sup>100</sup> Vgl. Rupprecht-Däullary (1994): 23f.

<sup>101</sup> Vgl. die Definition in Abschnitt 2.1.

<sup>102</sup> Vgl. Rupprecht-Däullary (1994): 18; sowie die dort angegebenen Quellen.

abzugrenzen,<sup>103</sup> die allerdings durchaus identische Ziele verfolgen kann.<sup>104</sup> In der Umsetzung kann wieder differenziert werden nach einer Eingliederung der Kooperation in den regulären Geschäftsbetrieb und der Bildung einer eigenständigen, dritten Unternehmung.<sup>105</sup> Die häufigste Form der Gemeinschaftsunternehmen mit der Bildung einer separaten Geschäftseinheit stellt dabei das *Joint Venture* (JV) dar.<sup>106</sup> SCHWERK unterscheidet für diese – wie für andere Formen – dann noch mehrere Arten einer Vertragsbasis zur Definition der verfügbaren Rechte und Pflichten.<sup>107</sup> Die Überlegung dieser Ausgestaltung von (Un-)Abhängigkeit gemeinschaftlich geführter und/oder verwalteter Kooperationen wird insbesondere dann relevant, wenn geschäftskritische Daten ausgetauscht werden und nur ein (Teil der) Partner über die (physische) Hoheit von Unternehmensdaten verfügt.<sup>108</sup>

#### 2.2.4 Kooperationsintensität

Der Begriff der Intensität wird in der Literatur nicht so eindeutig gehandhabt, wie es auf Basis des vorgestellten morphologischen Kastens erwartbar wäre.<sup>109</sup> So weist BALLING explizit darauf hin, dass auch die Aspekte des Formalisierungsgrades oder allgemein der erhaltenen Selbständigkeit häufig als Ausprägungen von Intensität genannt werden.<sup>110</sup> Es wird dabei deutlich, dass die identifizierten Kriterien zwar die wesentlichen und immer wieder diskutierten sind, jedoch auch Abhängigkeiten zwischen ihnen bestehen und die begriffliche Verwendung nicht

<sup>103</sup> Vgl. Killich (2011): 19f.

<sup>104</sup> Dazu gehören z. B. die Ergänzung des Produktportfolios oder die Schaffung von Synergieeffekten (vgl. Abschnitt 2.2.2). Der Einsatz von Unternehmensübernahmen als Mittel zur Erweiterung der eigenen Geschäftstätigkeit hat dabei in den vergangenen Jahren erheblich zugenommen (vgl. hierzu die Ausführungen zu Mergers & Acquisitions(M&A)-Wellen bei Coenenberg/Salfeld (2007): 186–188 sowie die ergänzenden Daten von Thomson Reuters (2015)).

<sup>105</sup> Vgl. Balling (1998): 154.

<sup>106</sup> Vgl. Fontanari (1996): 43; Schwerk (2000): 44. Ein Joint Venture (wörtlich aus dem Englischen: ‚Gemeinsames Wagnis‘) ist eine „spezifische Kooperationsform; die Partnerunternehmen sind jeweils mit Kapital am Joint Venture beteiligt, tragen gemeinsam das finanzielle Risiko der Investition und nehmen Führungsfunktionen im gemeinsamen Unternehmen wahr.“ (Engelhard/Altmann (2015)). Es stellt damit eine Dreierbeziehung zwischen zwei Mutter- und einem Tochterunternehmen dar (vgl. Sydow/Duschek (2011): 109). Auf die Schwierigkeiten zur Bestimmung des korrekten Verständnisses und der Einordnung in den Kontext weiterer Kooperationsarten vgl. Abschnitt 2.3.1.

<sup>107</sup> Vgl. Schwerk (2000): 41–43.

<sup>108</sup> Vgl. hierzu als Vorgriff auf die weiteren Teile der Arbeit auch die Überlegungen zur Beibehaltung der Autonomie in den Abschnitten 2.4 und 4.1.

<sup>109</sup> Vgl. Abb. 7.

<sup>110</sup> Vgl. Balling (1998): 160.

immer klar ist. Diese Arbeit folgt im Weiteren den Aufteilungen von RUPPRECHT-DÄULLARY und OSIECKA.<sup>111</sup>

Eine *Absprachenkooperation* liegt vor, wenn die beteiligten Unternehmen gegenseitige Verpflichtungen formulieren, die weiterhin keiner Interaktion bedürfen. Dies ist z. B. der Fall, wenn bestimmte Liefer- oder Abnahmemengen definiert werden, Forschungsfelder festgelegt werden oder die Duldung bestimmter Praktiken des jeweils anderen vereinbart wird. Bei einer *Austauschkooperation* (oder *reziproken Kooperation*) hingegen, findet ein Geben und Nehmen der Partner untereinander statt. Dies kann sowohl materielle wie auch immaterielle Güter betreffen, dazu gehören auch Daten, Informationen und Wissen.<sup>112</sup> Als dritte Variante kann die *Gemeinschaftskooperation* (oder *redistributive Kooperation*) angesehen werden. In diesem Fall übernehmen die Beteiligten gemeinschaftlich Aufgaben und setzen dafür kombinierte Ressourcen ein. In dieser Variante bedarf die Aufteilung des Kooperationserfolgs oder Kooperationsergebnisses besonderer Beachtung, wenngleich dies auch generell ein zentrales Thema bei der tatsächlichen Umsetzung von Kooperationen darstellt.<sup>113</sup>

### 2.3 Ausgewählte Konzepte unternehmensübergreifender Kooperation

In Vorbereitung auf die Ausführungen zu unternehmensübergreifenden BI-Netzwerken werden nachfolgend zwei in der Literatur vielbeachtete Konzepte aus dem Bereich der Kooperation beschrieben. Sie sollen zum einen die betriebswirtschaftliche Problemstellung der Notwendigkeit einer Allianzbildung motivieren

---

<sup>111</sup> Vgl. hierzu und zum Folgenden Osiecka (2006): 10; Rupprecht-Däullary (1994): 21f.

<sup>112</sup> Die Unterscheidung von Daten, Informationen und Wissen wird in der vorliegenden Arbeit, insbesondere bei der Beschreibung der verwendeten Literatur, nicht so vollständig trennscharf vorgenommen, wie es die Begriffe im eigentlichen Sinne verlangen. Ohne auf umfangreiche und formalisierte Ausführungen wie z. B. bei KRCMAR einzugehen (vgl. Krcmar (2010): 15) wird hier das Verständnis von MERTENS ET AL. angewandt: *Daten* sind eine Folge von Zeichen, die reale Objekte und deren Beziehungen beschreiben; *Informationen* entstehen durch die zweckorientierte Auswertung dieser Daten; *Wissen* wiederum entsteht durch die (personalisierte) Interpretation von Informationen (vgl. Mertens et al. (2012): 38, 52, 57). Während z. B. für die Bestimmung der Qualität von Daten und Informationen zunächst andere Maßstäbe gelten können, erfolgt ihr Austausch im Kontext von Kooperationen kaum trennscharf, wie sich z. B. bei RUPPRECHT-DÄULLARY (vgl. Rupprecht-Däullary (1994): 22) zeigt. Es bleibt auch offen, ob bei einem Austausch tatsächlich Interpretationen mit übermittelt werden und ab welchem Punkt diese tatsächlich vorliegen und kein reiner Datenaustausch stattfindet. Diese Begriffsproblematik zeigt sich auch bei spezialisierten Ausführungen (vgl. Abschnitt 3.3).

<sup>113</sup> Vgl. Balling (1998): 155.

und zum anderen existierende Ideen zum strukturierten Informationsaustausch aufzeigen.

### 2.3.1 Strategische Allianzen

Strategische Allianzen werden nicht nur ausführlich in der Literatur behandelt, sie werden auch als maßgeblich wertschöpfend und wachstumsfördernd wahrgenommen.<sup>114</sup> Tatsächlich wird der Begriff aber sehr umfassend für diverse Formen und Arten von Kooperation gebraucht.<sup>115</sup> Einigkeit besteht im Wesentlichen darin, dass die Allianzen langfristig orientiert sind und strategische Ziele verfolgen, anstatt das operative Geschäft optimieren zu wollen.<sup>116</sup> Werden die Ausprägungen der Kriterien betrachtet, wie sie für strategische Allianzen zutreffen können, so besteht bereits bei dem wichtigsten Kriterium, der Kooperationsrichtung, Uneinigkeit. Während SYDOW/DUSCHEK sowohl vertikale, als auch horizontale und diagonale Kooperationen unter dem Begriff zusammenfassen,<sup>117</sup> zeigt MORSCHETT anhand einer Literaturbetrachtung, dass unter Allianzen zumeist horizontale Kooperationen verstanden werden, da von ähnlichen Geschäftsfeldern ausgegangen wird.<sup>118</sup> Für vertikale Kooperationen bringt er den Begriff des ‚strategischen Netzwerks‘ ein, der in einer anderen Arbeit von SYDOW wiederum als der übergeordnete Begriff zu strategischen Allianzen verstanden wird.<sup>119</sup>

Das bereits bei der Vorstellung unterschiedlicher Kooperationsformen<sup>120</sup> genannte Konzept des Joint Ventures findet sich in der Literatur mehrfach als eine spezielle, durchaus auch als die vorherrschende, Form einer strategischen Allianz. So unterscheidet IHRIG Lizenzverträge, Managementverträge und Joint Ventures als Formen der strategischen Allianz und auch SYDOW ordnet sie entsprechend ein.<sup>121</sup> GIBBS/HUMPHRIES andererseits beginnen ihre Ausführungen mit dem Vermerk:

<sup>114</sup> Vgl. Sydow/Duschek (2011): 104.

<sup>115</sup> Vgl. Sydow/Duschek (2011): 105. Bereits FONTANARI weist darauf hin, dass bei der Abgrenzung der strategischen Allianz zu anderen Kooperationsformen eine gewisse „Hilflosigkeit“ (Fontanari (1996): 81) vorliegt und eine Trennung nicht klar erkennbar ist.

<sup>116</sup> Vgl. Sell (2002): 24; Sydow/Duschek (2011): 105. Eine umfangreiche Literaturbetrachtung liefert Knoppe (1997): 38–44. Zu aktuellen Überlegungen zu ‚strategischen Partnern‘ in Allianzen am oft zitierten Beispiel der ‚StarAlliance‘ vgl. auch Gerlach/Cleophas/Kliwer (2013): 147.

<sup>117</sup> Vgl. Sydow/Duschek (2011): 104.

<sup>118</sup> Vgl. Morschett (2005): 391f.

<sup>119</sup> Vgl. Sydow (1992): 61.

<sup>120</sup> Vgl. Abschnitt 2.2.3.

<sup>121</sup> Vgl. Ihrig (1991): 29 (zit. nach Sydow/Duschek (2011): 105); Sydow (1992): 62.



„Strategic alliances should not be confused with joint ventures.“<sup>122</sup> Es soll an dieser Stelle keine definitorische Lösung gefunden werden. Vielmehr stellt sich vor dem Hintergrund der Diskussion dar, dass die Betrachtung einzelner Kriterien von Kooperationen und die Wahl der für die gegebene Situation angemessenen Ausprägungen eine klarere Aussage über die Art der Partnerschaft treffen kann, als ein ‚plakativer‘ Begriff. Unstrittig bleibt die Wichtigkeit der strategischen Kooperation, insbesondere für kleine und mittlere Unternehmen.<sup>123</sup>

### 2.3.2 Supply Chain Management

Die in den vorherigen Abschnitten bereits diskutierte Wertschöpfungskette wird im Supply Chain Management intensiv betrachtet. Es interessiert dabei zunächst grundsätzlich eine vertikal gerichtete Kooperation. Der Begriff des SCM erfährt seit etwa 25 bis 30 Jahren eine gehäufte Verwendung.<sup>124</sup> Eine umfassende Definition liefern KUHN/HELLINGRATH: „Supply Chain Management ist die integrierte prozessorientierte Planung und Steuerung der Waren-, Informations- und Geldflüsse entlang der gesamten Wertschöpfungskette vom Kunden bis zum Rohstofflieferanten [...]“<sup>125</sup>. Dabei werden nicht nur die logistischen Prozesse betrachtet, sondern auch die weiteren Primär- und Sekundäraktivitäten, wie Einkaufs- oder Forschungs- und Entwicklungsprozesse.<sup>126</sup> Die komplexe Verflechtung von spezialisierten Rohstoffproduzenten und weiterverarbeitenden Unternehmen führt dabei zu einer in der Praxis nicht rein linear betrachtbaren Kette weniger Unternehmen, sondern zu einem Netzwerk, wenngleich auch in diesem noch die vertikale Integrationsrichtung im Fokus steht.<sup>127</sup>

Für die enge Verzahnung von Unternehmen einer Wertschöpfungskette sind in den letzten Jahren mehrere Konzepte erarbeitet und in der Praxis erprobt worden, die sich in vielen Fällen auf die operativen und Warenfluss-orientierten Prozesse beziehen.<sup>128</sup> Eine Methode, die ein hohes Maß an Informationsaustausch und Vertrauen verlangt, ist die des *Vendor-Managed Inventory* (VMI). Für den Konsum-

---

<sup>122</sup> Gibbs/Humphries (2009): 50.

<sup>123</sup> Vgl. Camarinha-Matos et al. (2009): 46; Doz (1992): 61; sowie zu höheren Synergieeffekten bei KMU: Paprotka (1996): 120.

<sup>124</sup> Vgl. Aliche (2005): 3; Lindemann (2005): 283.

<sup>125</sup> Kuhn/Hellingrath (2002): 10. Der ausgelassene Teil des Zitats erläutert die Ziele des SCM.

<sup>126</sup> Vgl. Cooper/Lambert/Pagh (1997): 10.

<sup>127</sup> Vgl. Gibbs/Humphries (2009): 48f.; Kurbel (2013): 222–224.

<sup>128</sup> Vgl. Dudek (2009): 239.



güterbereich entwickelt, sieht VMI die konstante Überprüfung und Verwaltung der Lagerbestände in der Kette nachgelagerter Unternehmen vor.<sup>129</sup> Stärker an gemeinschaftlicher Vorausplanung und -abstimmung orientiert ist das *Collaborative Planning, Forecasting & Replenishment* (CPFR). Hier werden auf grobgranularer Ebene Angebotsprogramme und entsprechende Lieferstrecken abgesprochen.<sup>130</sup> Das diesem begrifflich innewohnende *Collaborative Planning* grenzt STADTLER noch ab und beschreibt es als einen Prozess zur gemeinschaftlichen Findung eines konkreten operativen Plans, während CPFR ein auf Informationsaustausch fokussiertes, allgemeines Vorgehen darstellt.<sup>131</sup> Für den Bereich des SCM, bzw. der vertikalen Kooperation, existieren somit erfolgreich eingesetzte zwischenbetriebliche Informationsaustauschkonzepte, die helfen, Fehlplanungen und Falschinterpretationen von Geschäftsereignissen zu vermeiden.<sup>132</sup>

## **2.4 Fallbeispiel zur Kooperation in Analyseprozessen: Auswertung der gemeinsamen Produktentwicklung in Zulieferernetzwerken**

Die bisherigen Ausführungen zeigen, dass Kooperationen zwischen Unternehmen einen hohen Gestaltungsspielraum in vielen Kriterien aufweisen. Im Folgenden wird eine Konstellation aufgezeigt und begründet, die für die weitere Arbeit die gedankliche Grundlage bilden und die betriebswirtschaftliche Motivation verdeutlichen soll.

Obwohl ein Großteil der beschriebenen Kooperationssysteme auf die primären Wertaktivitäten Bezug nimmt und insbesondere Dispositions- und Produktionsprozesse fokussiert, zeigt sich, dass bei in der Praxis bestehenden Kooperationen von Unternehmen *Forschung und Entwicklung* (F&E) die mit Abstand am Häufigsten für Kooperation genutzte Unternehmensfunktion darstellt. In ihrer Studie mit 30.000 angeschriebenen Unternehmen unterschiedlicher Branchen aus dem Gebiet der Bundesrepublik Deutschland führen HAUSCHILD/WALLACHER eine

---

<sup>129</sup> Vgl. Kurbel (2013): 230.

<sup>130</sup> Vgl. Dudek (2009): 240; Kurbel (2013): 231.

<sup>131</sup> Vgl. Stadler (2009): 6, 10f.

<sup>132</sup> Als Argumentationsbeispiel wird typischerweise der ‚Bullwhip-Effekt‘ (auch Peitscheneffekt) herangezogen. Bei einer mehrstufigen Produktion führt eine geringe Erhöhung des Bedarf am Ende zu einer durch Puffer und Überinterpretation immer weiter ausgebauten Produktion in vorgelagerten Stufen, für die keine tatsächliche Nachfrage besteht, sodass Lagerkosten und ggf. Kosten durch Vernichtung von Überproduktion entstehen. Weitere Effekte mit ähnlicher Argumentationsbasis sind der ‚Forrester-Effekt‘ und der ‚Burbidge-Effekt‘ (vgl. Alicke (2005): 99–109; Kuhn/Hellingrath (2002): 17–20).

Nutzungsquote von 25,0% an, gefolgt von ‚Information und Kommunikation‘ mit 20,9%.<sup>133</sup> Als Motivation nennen die Unternehmen dabei für den Bereich F&E hauptsächlich ‚Wissensaustausch zum optimalen Einsatz der eigenen Ressourcen‘ (77,7% Nennungen mit der Beurteilung ‚wichtig‘ oder ‚sehr wichtig‘).<sup>134</sup> So verweist auch PAPROTTKA bereits auf den positiven Effekt gemeinsamer F&E-Aktivitäten, der nicht nur im F&E-Bereich selbst, sondern, z. B. durch das Design einheitlicher Schnittstellen der entwickelten Produkte, auch im folgenden Produktions- und späteren Wartungsprozess Synergiepotenziale entstehen lässt.<sup>135</sup> Diese Überlegungen und Ergebnisse deuten darauf hin, dass hier ein Bedürfnis nach einer Austauschkooperation besteht, die weniger auf den direkten Güter- als vielmehr auf den Informationsfluss innerhalb von Wertschöpfungsnetzwerken fokussiert. Dies legt die Betrachtung der unternehmerischen Tätigkeit in nicht-produzierenden Aktivitäten nahe. Dazu gehören auch, aber nicht nur, die Führungsfunktionen einer Unternehmung.<sup>136</sup>

Zur Verdeutlichung soll die folgende Einordnung von Unternehmensaktivitäten dienen. Ein Unternehmen kann als ein System verstanden werden, das mit Planwerten zur Steuerung versehen werden kann. Diesen Planwerten stehen Ist-Werte entgegen. Die Prüfung der Abweichung von Plan und Ist sowie die daraus abzuleitenden Handlungen können in einem Regelkreis beschrieben werden. Dieser bezieht dabei sowohl ein steuerndes, in die Zukunft gerichtetes Instrumentarium (‚feed forward‘) als auch ein kontrollierendes, Abweichungen verdeutlichendes (‚feed back‘) mit ein.<sup>137</sup> Das generische Modell von HUCH/BEHME/OHLENDORF

---

<sup>133</sup> Vgl. Hauschild/Wallacher (2004): 1012. HAUSCHILD/WALLACHER geben in ihrer Erhebung die Anzahl der angeschriebenen Unternehmen mit 30.000 an, nennen jedoch nicht die Rücklaufquote oder die Zahl der Antworten (vgl. Hauschild/Wallacher (2004): 1009f.). Für niedersächsische Unternehmen, die an dieser Studie beteiligt waren, wird in einer anderen Publikation eine Rücklaufquote von ca. 32,4% angegeben (940 von 2900) (vgl. Volkmann (2004): 350). Unter der Annahme, dass diese Quote auch in den anderen Bundesländern in ähnlicher Höhe erzielt wurde, kann von einer Stichprobengröße von fast 10.000 Unternehmen ausgegangen werden ( $30.000 \times 32,4\% \approx 9.720$ ).

<sup>134</sup> Vgl. Hauschild/Wallacher (2004): 1012.

<sup>135</sup> Vgl. Paprottko (1996): 84f. Entsprechende Erkenntnisse ergeben sich auch aus der Auswertung der Fördermaßnahmen der Regierung der Bundesrepublik Deutschland, die bereits seit den 1970er-Jahren gezielt die F&E kleiner und mittlerer Unternehmen stärken sollen (vgl. Wolff et al. (1994): 1–42).

<sup>136</sup> Vgl. hierzu die Ausführungen im einführenden Text zu diesem Kapitel.

<sup>137</sup> Vgl. Huch/Behme/Ohlendorf (2004): 228f.

überführen GANSOR/TOTOK/STOCK in einen Regelkreis der Unternehmenssteuerung, der die Aspekte der Entscheidungsunterstützung betont (vgl. Abb. 8).<sup>138</sup>

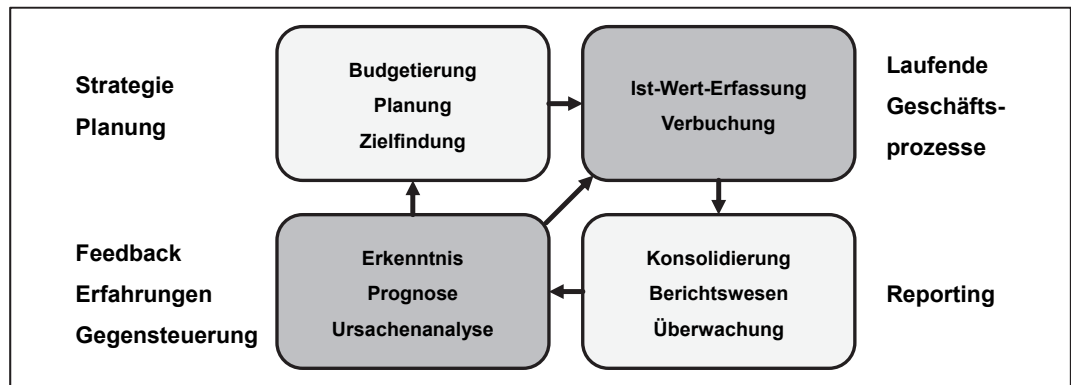


Abb. 8: Regelkreis der Unternehmenssteuerung

(Quelle: In Anlehnung an Gansor/Totok/Stock (2010): 45)

Analytische Prozesse finden dabei insbesondere in den Phasen ‚Reporting‘ und ‚Feedback, Erfahrungen, Gegensteuerung‘ statt. Die hier gewonnenen Erkenntnisse sind zunächst unternehmensintern und -begrenzt. Ein Austausch dieser Informationen bedingt zwangsläufig Überlegungen zur Wettbewerbsrelevanz der Daten und zu deren potenziellem Missbrauch durch Partnerunternehmen, die ggf. gleichzeitig im Wettbewerb stehen, was insbesondere bei kleineren, hochspezialisierten Unternehmen und horizontalen Kooperationsformen zum Tragen kommt. Dennoch sind KMU mehr als andere in der Lage, durch Kooperationen Synergieeffekte zu realisieren.<sup>139</sup>

Mit 99,3% aller Unternehmen stellen KMU die bei weitem überwiegende Anzahl an Unternehmen in Deutschland dar, die für nahezu die Hälfte der Bruttowertschöpfung verantwortlich zeichnet.<sup>140</sup> Es besteht daher nicht nur aufgrund des erwarteten Mehrwertes eines einzelnen Unternehmens, sondern auch aus gesamtwirtschaftlicher Sicht ein Interesse daran, die Kooperationsfähigkeit von KMU zu fördern, zumal diese einen entscheidenden Faktor für die Überwindung kleingrö-

<sup>138</sup> Ähnliche Ausführungen zu dem Zusammenspiel der genannten vier Phasen liefern auch HAHN/HUNGENBERG (vgl. Hahn/Hungenberg (2001): 46). Als generischer Ausgangspunkt dieser vier-phasigen Verbesserungszyklen wird üblicherweise die Arbeit von DEMING genannt. Er formuliert einen Kreislauf (vgl. Deming (2000): 88), den er auf SHEWHART zurückführt und der üblicherweise als PDCA-Zyklus bezeichnet wird, wobei PDCA die vier Phasen ‚Plan‘, ‚Do‘, ‚Check‘ und ‚Act‘ beschreibt (vgl. Hummeltenberg (2010): 23; Koch (2011): 118f.).

<sup>139</sup> Vgl. Abschnitt 2.2.2.

<sup>140</sup> Vgl. Statistisches Bundesamt (2013): 504.

Benbedingter Limitierungen in der Ressourcenverfügbarkeit darstellt.<sup>141</sup> Dazu können insbesondere horizontale Kooperationsformen betrachtet werden, bei denen mehrere Unternehmen ähnliche Kompetenzen bündeln und so eine ‚kritische Masse‘, z. B. in der Entwicklung von Technologien, erreichen. Eine mögliche Ausprägung davon beschreiben TEICH/FISCHER/KÄSCHEL mit der Darstellung von ‚Competence Cells‘ als nicht-hierarchisch organisierten Netzwerken.<sup>142</sup> Im Zuge dieser Kooperationen entstehen Verflechtungen mehrerer Dutzend Unternehmen, die ihre Unabhängigkeit wertschätzen und zeitlich flexibel mit ihren Partnern kooperieren, bzw. Kooperationen auch unproblematisch zu beenden suchen, wenn kein Vorteil mehr erzielbar ist, was insbesondere auf Seiten der Partner immer in die Überlegungen bei der Schaffung eventueller Abhängigkeiten mit einbezogen werden muss.<sup>143</sup>

Unterschiede in der Kooperationsbereitschaft zeigen sich dabei sowohl regional als auch branchenabhängig. So weisen Studien darauf hin, dass Unternehmen in den ostdeutschen Bundesländern eine deutlich höhere und in den vergangenen Jahren auch deutlich gestiegene Bereitschaft aufweisen.<sup>144</sup> Branchenspezifisch lassen sich die relativ höchsten Werte von Unternehmen mit Kooperationen für den Bereich Dienstleistungen (42,7%) sowie das verarbeitende Gewerbe (32,7%) erkennen.<sup>145</sup> Diese Werte sind, was auch dem zeitlichen Abstand der Studien geschuldet sein mag, in aktuellen Erhebungen in der Automobilindustrie deutlich höher. LANGE zeigt, dass 84% aller befragten Unternehmen dieser Branche in mindestens zwei Kooperationen zum Zeitpunkt der Befragung aktiv waren (Stand 2011).<sup>146</sup> Das Ergebnis der Studie ähnelt dabei auch dem der empirischen Erhebung von MICHEL, der beschreibt, dass etwa 74% zu Kooperationen im Bereich F&E befragter Unternehmen mindestens zwei aktive Kooperationen aufweisen.<sup>147</sup> Die Automobilindustrie ist daher nicht nur wegen ihrer wirtschaftlichen Bedeutung,<sup>148</sup> sondern auch wegen ihrer Kooperationsfreudigkeit ein vielzitiertes Bei-

---

<sup>141</sup> Vgl. Camarinha-Matos et al. (2009): 46.

<sup>142</sup> Vgl. Teich/Fischer/Käschel (2002): 178f.

<sup>143</sup> Vgl. Mettler/Pinto/Raber (2012): 1.

<sup>144</sup> Vgl. Schmidt/Kiefer (2005): 1369; sowie die dort aufgeführte Übersicht von Studien.

<sup>145</sup> Vgl. Hauschild/Wallacher (2004): 1012.

<sup>146</sup> Vgl. Lange (2011): 11.

<sup>147</sup> Vgl. Michel (2009): 86.

<sup>148</sup> Mit über 375 Mrd. EUR Umsatz (ca. 6,4% des Gesamtumsatzes der deutschen Wirtschaft 2011) stellt die Automobilbranche die größte Einzelbranche Deutschlands dar (vgl. Statistisches Bundesamt (2013): 500).

spiel für Zusammenarbeit von Unternehmen.<sup>149</sup> Dabei stellt sich heraus, dass für das Eingehen einer Kooperation nur die Aspekte ‚Know-How‘ und ‚Kernkompetenzen‘ der Partner mehr wertgeschätzt werden als das ‚Vertrauen‘ in die Partnerschaft.<sup>150</sup> Liegt dieses vor, so wird die Gefahr eines Know-How-Abflusses auch als beherrschbar eingestuft.<sup>151</sup> Neben den allgemeinen Studien existieren auch konkrete Umsetzungen dieser Netzwerke. Als Beispiel wird hier das ‚Netzwerk Automobilzulieferer Sachsen‘ (AMZ) aufgeführt, dessen Schwerpunkt die Förderung der Kommunikation zwischen den verbundenen Unternehmen darstellt. Es besteht dabei keinerlei formale Abhängigkeit der Unternehmen, sondern vielmehr ein freiwilliger und projektbezogener Austausch.<sup>152</sup>

Im Rahmen dieser Arbeit soll das beschriebene Netzwerk als realer Ideengeber für ein Fallbeispiel herangezogen werden, sodass eine betriebswirtschaftliche Motivation für die weiteren Ausführungen gegeben ist.<sup>153</sup> Angenommen wird der Fall, dass ein Netzwerk unabhängiger KMU aus der Automobilindustrie im Bereich F&E zusammenarbeiten will. Neben dem ggf. notwendigen Güterfluss ist insbesondere ein Informationsfluss zwischen den Unternehmen vonnöten. Im Rahmen der Entscheidungsunterstützung<sup>154</sup>, welche F&E-Projekte mit welchen Partnern vielversprechend erscheinen, sollen Kennzahlen bereits durchgeführter Projekte verglichen werden. Dazu wird ein einfaches Modell erstellt, das die Strukturierung der Kennzahlen ‚Umsatz‘, ‚Investition‘ und ‚Anzahl Projektpartner‘ nach den Merkmalen ‚Produkt(kategorie)‘, ‚Absatzmarkt für die Entwicklung‘ und ‚Zeitpunkt/-raum der Entwicklung‘ ermöglicht. Die konkrete Abbildung und Ausgestaltung dieses Modells sind Gegenstand von Abschnitt 6.2.

Das vorgestellte Modell ist vergleichsweise einfach strukturiert und kann vor dem diskutierten betriebswirtschaftlichen Hintergrund auch komplexer aufgebaut wer-

---

<sup>149</sup> So listet z. B. LANGE 20 Studien mit Kooperationsschwerpunkt für die Automobilindustrie im Zeitraum 2001–2009 auf (vgl. Lange (2010): 26f.).

<sup>150</sup> Auf einer Skala von 0 (unwichtig) bis 10 (sehr wichtig) wird das ‚Vertrauenspotenzial‘ im Mittel mit 8,23 (‚Know-How‘: 8,73; Kernkompetenzen: 8,56) eingeschätzt und liegt damit z.T. deutlich vor z. B. ‚Persönlicher Kontakt‘ (6,98) und ‚Kooperationserfahrung‘ (5,23) (vgl. Michel (2009): 95).

<sup>151</sup> Vgl. Michel (2009): 105.

<sup>152</sup> Das AMZ wurde von SCHOLTA im Rahmen einer Dissertation umfänglich untersucht, die beschriebenen Eigenschaften dabei besonders herausgestellt (vgl. Scholta (2005): 30f.).

<sup>153</sup> Generell ist die Automobilindustrie immer wieder eine in der wissenschaftlichen Literatur herangezogene Branche für Modelle und Untersuchungen (vgl. zu den bereits angegebenen Quellen auch Göltenboth (1998): 14–36).

<sup>154</sup> Vgl. Kapitel 3.

den. Für die im Rahmen dieser Arbeit zu diskutierenden Forschungsfragen ist jedoch weniger die umfangreiche betriebswirtschaftliche Struktur des Netzwerks als vielmehr die Analyse von Aspekten der Entscheidungsunterstützung im Kontext von Netzwerken relevant. Das Fallbeispiel liefert damit zunächst eine Grundlage und Motivation der diskutierten Konzepte nachfolgender Kapitel.<sup>155</sup>

In Abb. 9 wird abschließend der erarbeitete morphologische Kasten verwendet, um die Struktur des im Fallbeispiel verwendeten Netzwerks in bekanntem Kontext darzustellen. Die zutreffenden Ausprägungen der Kriterien sind jeweils grau hinterlegt dargestellt. Dabei sind die markantesten fünf dunkelgrau dargestellt. So ist z. B. das Vorliegen einer ‚projektbezogenen‘ Fristigkeit von größerer Wichtigkeit als die Festlegung des Raumaspektes, da hier durch die reine Verwendung von Informationstechnologie keinerlei physische Nähe vorliegt.<sup>156</sup>

Kriterien	Ausprägungen			
Richtung / Wirtschaftsstufe	Horizontal	Vertikal	Diagonal / Lateral / Konglomerat	
Bindung / Vertragsform	Formal	Informell	Stillschweigend	Gewohnheitsrechtlich
	Kodifiziert		Nicht kodifiziert	
Form / Beteiligungsverhältnis	Lizenzvereinbarung	Managementvertrag	Franchising	Joint Venture
	Minderheit	Mehrheit	Parität	Keine Beteiligung
Fristigkeit / Dauer / Stabilität	Einmalig	Sporadisch	Regelmäßig	Dauerhaft
	Projektbezogen		Projektübergreifend	
	Kurzfristig	Mittelfristig	Langfristig	Unbefristet
Intensität / Art der Verflechtung	Absprachenkooperation		Austauschkooperation	
	Gemeinschaftskooperation		Gemeinschaftskooperation	
Unternehmensbereiche / Funktionen	Austausch von Informationen und Ereignissen		Koordiniert arbeitsteiliges Vorgehen	
	Primäraktivitäten		Sekundäraktivitäten	
Anzahl der Partner / Koordinationsstruktur	F&E	Beschaffung	Produktion	Weitere Funktionen
	Bilateral	Trilateral	Einfache Netzwerke	Komplexe Netzwerke
	Niedrig		Hoch	
Raumaspekt / Partnerherkunft	Lokal	Regional	National	International
Zielsysteme / Zielkompatibilität	Identisch	Komplementär		Neutral
	Hoch		Niedrig	

Abb. 9: Ausprägungen der Kooperationskriterien im Fallbeispiel

(Quelle: Eigene Darstellung)

<sup>155</sup> Komplexere Anwendungsfälle werden z. B. im Rahmen des Projektes ‚DIPRO-BI‘ umgesetzt, das den Aufbau einer produktorientierten Datenbank beschreibt, um u. a. ein übergreifendes Fehlermanagement und entsprechende Analysen oder Einsparpotenzialberechnungen über alle beteiligten Prozesspartner hinweg zu ermöglichen (vgl. Kemper et al. (2014): 24–30).

<sup>156</sup> Zum Aufbau des morphologischen Kastens vgl. Abschnitt 2.2.1. An dieser Stelle sollen rechtliche Aspekte (Datenschutz, etc.) und technische Aspekte (Unterschiedliche Sprachversionen von Systemen, etc.) bei der Überschreitung von Ländergrenzen zunächst nicht behandelt werden, da sie für die konzeptionelle Struktur des Netzes im Rahmen dieser Arbeit eine untergeordnete Rolle spielen, während z. B. die Anzahl der beteiligten Partner einen erheblichen Einfluss auf die Gestaltung eines Netzes als solchem hat (vgl. hierzu auch die Ausführungen zur Netzgestaltung in den Kapiteln 4 und 5).

### 3 Entscheidungsunterstützung in Unternehmen

Das Konzept der Entscheidung wird bereits seit mehreren Jahrzehnten durch die Betriebswirtschaft betrachtet und dort als zentraler Begriff behandelt, sodass sich die Betriebswirtschaftslehre zumindest in Teilen als angewandte Entscheidungslehre betrachten lässt.<sup>157</sup> Der Begriff der Entscheidung wird dabei im Grundverständnis auch heute noch so aufgefasst, wie ihn THOMAE bereits 1960 auf Basis einer Literaturanalyse beschreibt: „Entscheidung ist offensichtlich eine der Formen, in welcher der Mensch auf eine multivalente, d. h. also mehrere Möglichkeiten enthaltende und in mehrfacher Richtung auffordernde Situation reagiert“.<sup>158 159</sup>

Bei der Untersuchung menschlicher und organisatorischer Entscheidungen wird unterschieden zwischen einem präskriptiven Ansatz, dessen Ziel es ist, komplexe Entscheidungen zu unterstützen, und einem deskriptiven Ansatz, der existierendes Entscheidungsverhalten erklären soll.<sup>160</sup> Die in dieser Arbeit beschriebenen Ansätze und Systeme sind präskriptiv-orientiert zu verstehen. Ihr Ziel es ist, die Teilprozesse des Entscheidungsprozesses – Zielbildungsprozess, Informationsprozess und Auswahlprozess –<sup>161</sup> mit Hilfe von Informationstechnologie zu unterstützen.

In diesem Kapitel werden nach einer Einführung und Abgrenzung von Business Intelligence – als aktuellem Begriff für IT-gestützte Entscheidungsunterstützung – Gruppenentscheidungsprozesse und dafür entwickelte Systeme vorgestellt. Im Anschluss werden die unter dem Begriff Collaborative Business Intelligence subsummierten Ansätze und Werkzeuge zur gemeinschaftlichen und/oder gruppenorientierten Nutzung von BI-Systemen klassifiziert und erläutert.

#### 3.1 Business Intelligence

Der Begriff Business Intelligence wird in seinem heutigen Verständnis seit Anfang der 1990er-Jahre verwendet.<sup>162</sup> Er beschreibt IT-Systeme zur Management-

---

<sup>157</sup> Vgl. Kirsch (1971): 7.

<sup>158</sup> Thomae (1960): 18.

<sup>159</sup> Im Kontext der Betriebswirtschaftslehre wird in der Regel ergänzt, dass die Entscheidung ‚optimal‘ sein soll im Sinne einer Handlung, die die bestmöglichen Konsequenzen schafft (vgl. Erdmann/Popp (2006): 25f.).

<sup>160</sup> Vgl. Eisenführ/Weber (1999): 2f.

<sup>161</sup> Vgl. Bamberg/Coenenberg/Krapp (2008): 9.

<sup>162</sup> Vgl. Watson/Wixom (2007): 96. Die Nennung durch Analysten von Gartner, Inc. wird allgemein als Beginn der modernen Verwendung des Begriff angesehen, obwohl LUHN bereits 1958 die Grundidee eines Informationen sammelnden und verarbeitenden Systems als ‚Business Intelligence System‘ veröffentlichte (vgl. Luhn (1958): 314–319).



unterstützung, insbesondere zur Entscheidungsunterstützung, folgt jedoch trotz oder gerade wegen des weitreichenden Einsatzes der Methoden, Konzepte und Systeme keiner einheitlichen Definition.<sup>163</sup>

Im Folgenden wird BI daher zunächst vor dem Hintergrund der Entwicklung von Decision-Support-Systemen<sup>164</sup> eingeordnet und erklärt. Darauf aufbauend werden einzelne Kernbestandteile von BI-Systemen vertiefend diskutiert, soweit sie für das weitere Verständnis der Arbeit notwendig sind.

### 3.1.1 Definition und Einordnung in den Kontext von Decision-Support-Systemen

Der Zweck eines Decision-Support-Systems ist die interaktive und problembezogene Unterstützung eines Entscheidungsträgers mit Hilfe eines IT-Systems. Zu diesem Zweck werden geeignete Daten, Modelle und Methoden bereitgestellt.<sup>165</sup> Dabei sind charakteristische Merkmale die Verwendung direkt durch den Entscheider sowie die funktionale Fokussierung des Systems auf die jeweils unterstützte Aufgabe oder Aufgabenklasse.<sup>166</sup>

Die Anfänge der Entwicklung von DSS werden dabei von POWER auf die Mitte der 1960er-Jahre datiert.<sup>167</sup> Der Entwicklungspfad bis heute ist allerdings nicht einheitlich bestimmbar, was sich an den unterschiedlichen Klassifikationsmöglichkeiten der DSS im Laufe der Zeit zeigt. So unterscheidet ALTER 1980 sieben verschiedene DSS-Typen:<sup>168</sup>

- *File drawer systems* (Datenzugriff)
- *Data analysis systems* (Datenveränderungsmöglichkeiten)
- *Analysis information systems* (Zugriff auf Datenbanken und Modelle)
- *Accounting models* (Planungsrechnung basierend auf Konten)
- *Representational models* (Modell-basierte Rechnungen)

<sup>163</sup> Kemper/Baars/Mehanna (2010): 2f.

<sup>164</sup> Die Wahl einer einheitlichen Terminologie wird durch die Begriffsdupplung im Englischen und Deutschen noch erschwert. Während für ‚Business Intelligence‘ keine verbreitete deutsche Übersetzung existiert, wird auch im Deutschen für Entscheidungsunterstützungssysteme zu meist der englische Begriff Decision Support Systems oder die eingedeutschte Variante Decision-Support-Systeme (DSS) verwendet (vgl. Kemper/Baars/Mehanna (2010): 111; Hummeltenberg (2010): 19). Diese Arbeit verwendet daher im Folgenden die Bezeichnung DSS anstelle von ‚Entscheidungsunterstützungssystem(en)‘.

<sup>165</sup> Vgl. Gluchowski/Dittmar/Gabriel (2008): 63.

<sup>166</sup> Vgl. Mertens/Meier (2009): 12.

<sup>167</sup> Vgl. Power (2009).

<sup>168</sup> Vgl. hierzu und zum Folgenden Alter (1980): 74.



- *Optimization models* (Optimierung unter Restriktionen)
- *Suggestion models* (Entscheidungsvorbereitung strukturierter Aufgaben)

Diese Klassifikation weicht in Teilen von der moderneren Aufteilung durch POWER ab, was auch durch die technische Evolution im Entwicklungsprozess zu erklären ist.<sup>169</sup> Er unterscheidet

- *Model-driven DSS*, die Zugriff auf Finanz-, Optimierungs- und Simulationsmodelle bieten,
- *Data-driven DSS*, die Zugriff auf und Veränderung von Zeitreihen interner und externer Datenbestände fokussieren und dabei auch sehr große Bestände verwalten können,
- *Communications-driven DSS*, die in erster Linie Kommunikation und Kooperation der Beteiligten im Entscheidungsprozess stärken,
- *Document-driven DSS*, die auf die Verwaltung und Analyse von großen, ggf. multimedialen Dokumenten spezialisiert sind, sowie
- *Knowledge-driven DSS*, die im Sinne von Expertensystemen Entscheidungen aktiv durch Vorschläge herbeizuführen suchen.<sup>170</sup>

Es besteht in der Literatur weitgehend Einigkeit darin, dass BI im Rahmen der Data-driven DSS einzuordnen ist, da die für BI prägenden Methoden und Werkzeuge alle auf Datenerfassung, -aufbereitung und -präsentation beruhen.<sup>171</sup> Zu diesen gehören insbesondere Data Warehousing<sup>172</sup>, On-Line Analytical Processing<sup>173</sup> sowie ein einfacher, durch grafische Oberflächen getriebener Zugang zu aufbereiteten Daten, wie er in Executive Information Systems (EIS) oder Executive Support Systems (ESS) verwendet wird. Dieser Auffassung entsprechend sind auch die Visualisierungen der DSS-Zeitlinie (vgl. Abb. 10) und der BI-Evolution (vgl. Abb. 11) gestaltet. Die beiden nachstehenden Abbildungen stellen wesentli-

<sup>169</sup> Die Klassifikation von ‚Communication-driven DSS‘ mit dem Einsatz von Videotechnologie war 1980 z. B. nicht naheliegend (vgl. Power (2009)).

<sup>170</sup> Vgl. Power (2009). POWER geht bei Communication-driven DSS auch auf Gruppenentscheidungsunterstützungssysteme ein (vgl. dazu Abschnitt 3.2).

<sup>171</sup> Vgl. Hummeltenberg (2010): 26–31; Power (2009); Turban/Sharda/Delen (2011): 19f. Anzu-merken ist, dass z. B. LAUDON/LAUDON dieser Aufteilung nicht vollständig folgen und DSS als Spezialform von BI auffassen (vgl. Laudon/Laudon (2006): 459), obwohl sie im Weiteren data-driven DSS mit den gleichen Merkmalen beschreiben und On-Line Analytical Processing sowie Data Warehousing als Bestandteile definieren (vgl. Laudon/Laudon (2006): 466, 473).

<sup>172</sup> Vgl. hierzu Abschnitt 3.1.2.

<sup>173</sup> Vgl. hierzu Abschnitt 3.1.3.

chen Methoden, Konzepte und/oder Komponenten sowie ihren Einfluss aufeinander, bzw. ihre zeitliche Abhängigkeit dar.<sup>174</sup>

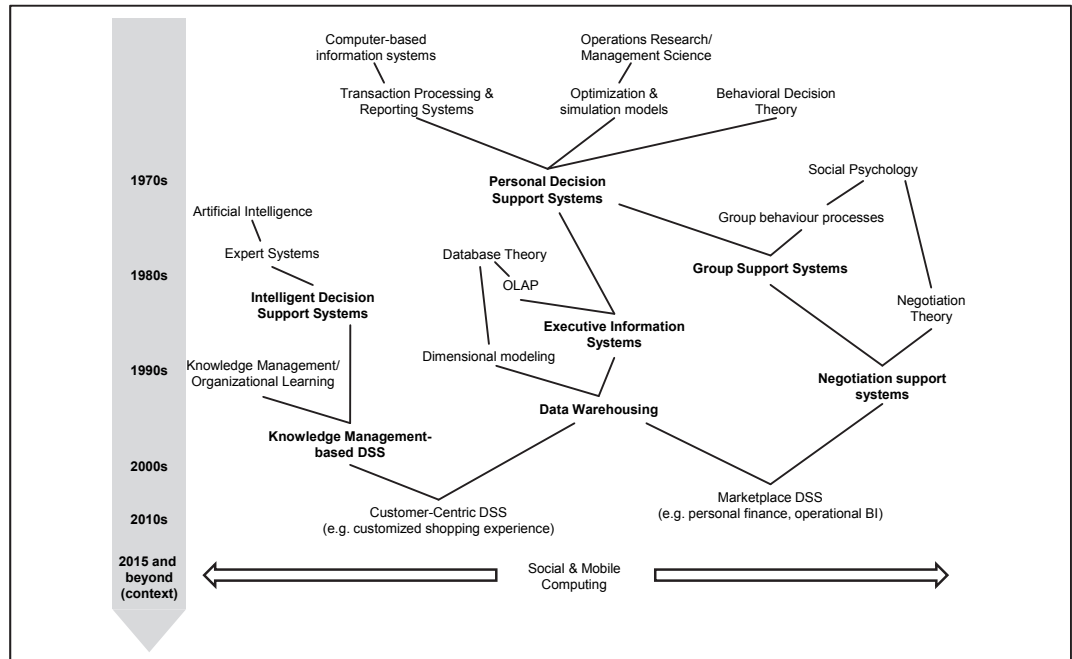


Abb. 10: DSS-Zeitlinie

(Quelle: In Anlehnung an Hosack et al. (2012): 328)

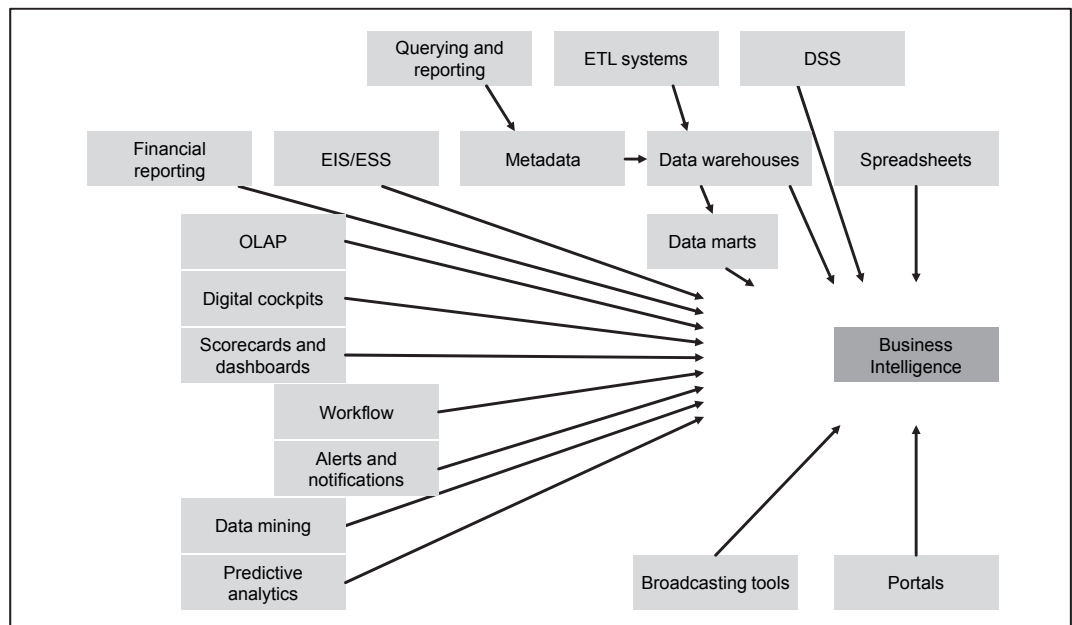


Abb. 11: BI-Evolution

(Quelle: In Anlehnung an Turban/Sharda/Delen (2011): 19)

<sup>174</sup> Die in den Abbildungen gezeigten Elemente und Verbindungen vermitteln eine Vorstellung des betrachteten Spektrums und Kontexts sowie der gegenseitigen Einflussnahme der Ansätze im Bereich DSS. Die für diese Arbeit maßgeblichen werden im Folgenden vorgestellt. Für eine detaillierte Einzelbetrachtung aller gezeigten Inhalte sei auf die jeweiligen Quellen verwiesen.

TURBAN/SHARDA/DELEN nutzen die Datenzentrierung der BI-Systeme als wichtigstes Unterscheidungsmerkmal eines BI-Systems zu einem ‚generischen‘ DSS: „[...] BI implies the use of a data warehouse, whereas DSS may or may not have such a feature.“<sup>175</sup> Die im deutschsprachigen Raum häufig zitierte Einteilung der Facetten von BI nach GLUCHOWSKI weist Data Warehousing hingegen nur im Rahmen des ‚weiten BI-Verständnisses‘ aus (vgl. Abb. 12).<sup>176</sup>

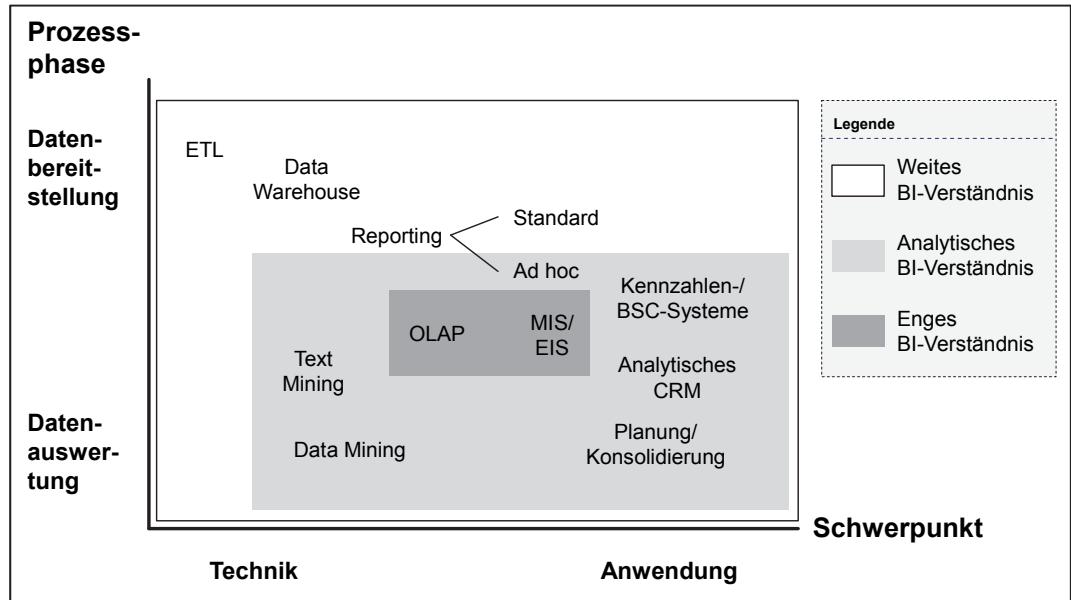


Abb. 12: Einordnung unterschiedlicher Facetten von Business Intelligence  
(Quelle: Gluchowski (2001): 7)

Die vorliegende Arbeit orientiert sich daher an der umfassenden Definition von KEMPER/BAARS/MEHANNA, die den gesamten, oben beschriebenen Umfang an Methoden, Werkzeugen und Systemen abdeckt: „Business Intelligence (BI) bezeichnet einen integrierten, unternehmensspezifischen, IT-basierten Gesamtansatz zur betrieblichen Entscheidungsunterstützung.“<sup>177</sup> Diese ‚weite‘ Definition ent-

<sup>175</sup> Turban/Sharda/Delen (2011): 23.

<sup>176</sup> Der hier vorgestellte Ordnungsrahmen zeigt auf der vertikalen Achse eine auf die Speicherung und Bereitstellung von Daten fokussierte Ebene (oben) und geht über zu einer methodisch geprägten Ebene (unten), bei der die nachfolgende Analyse der gespeicherten Daten im Vordergrund steht. GLUCHOWSKI trennt weiterhin zwischen einem technischen (links) und anwendungsorientierten (rechts) Schwerpunkt. Während die technische Betrachtung algorithmisch und systemisch getrieben ist (wie insbesondere bei Data Warehouse und dem verbundenen Extraktions-, Transformations- und Ladeprozess) finden sich auf der Anwendungsseite auch stark betriebswirtschaftlich motivierte Konzepte wie z. B. die Balanced Scorecard (BSC) (vgl. Gluchowski (2001): 8).

<sup>177</sup> Kemper/Baars/Mehanna (2010): 9. Im Original ist der Satzteil „integrierten ... Gesamtansatz“ hervorgehoben.

spricht auch dem aktuell vorherrschenden Verständnis in der Literatur,<sup>178</sup> wenn-  
gleich im Kontext des Regelkreises der Unternehmenssteuerung eine Fokussie-  
rung auf die retrospektiv orientierten Bereiche erkennbar ist.<sup>179</sup>

### 3.1.2 Data Warehousing

In der Literatur wird weitgehend übereinstimmend ein Data Warehouse (DWH) als eine unternehmensweit verfügbare Datenbasis angesehen, die losgelöst von den operativen Systemen eines Unternehmens betrieben wird.<sup>180</sup> In Bezug auf die technische Umsetzung können mehrere Varianten unterschieden werden,<sup>181</sup> wobei die Kernfrage darin besteht, ob eine zentrale Datenbank vorgehalten werden soll oder ob die Summe aller Datenbanken, die relevante Informationen vorhalten, als DWH angesehen werden soll.<sup>182</sup> Diese Arbeit folgt sowohl in der Definition der wesentlichen Merkmale als auch der technischen Grundstruktur der Variante von INMON, wie sie unter anderem auch von DEVLIN, HANSEN/NEUMANN und MUCKSCH/BEHME wiedergegeben wird: „A data warehouse is a subject-oriented, integrated, nonvolatile, and time-variant collection of data in support of management’s decisions.“<sup>183</sup>

Dabei meint

- *subject-oriented* eine Ausrichtung und Strukturierung der Daten anhand der sachlich in einem Unternehmen existierenden Sichtweisen (wie z. B. Produkte, Regionen) anstelle einer rein an den Funktionen (Einkauf, Produktion, Verkauf, etc.) angelehnten,

<sup>178</sup> Vgl. Chaudhuri/Dayal/Narasayya (2011): 90; Chen/Chiang/Storey (2012): 1166; Hummeltenberg (2010): 25; Watson/Wixom (2007): 96; sowie unter dem Begriff ‚Analytische Informationssysteme‘: Chamoni/Gluchowski (2010): 5.

<sup>179</sup> Vgl. Abschnitt 2.4. Retrospektiv beschreibt hier ‚Reporting‘ und ‚Feedback, Erfahrungen, Gegensteuerung‘ (vgl. Abb. 8). Als datengetriebene Form eines DSS ist BI demnach nicht vor-  
dergründig auf die Planung oder die Erfassung laufender Geschäftsprozesse ausgelegt.

<sup>180</sup> Vgl. Hansen/Neumann (2009): 1017f.; Inmon (2005): 28; Kimball/Ross (2002): 7; Mucksch/Behme (2000): 14.

<sup>181</sup> Vgl. Kemper/Baars/Mehanna (2010): 22.

<sup>182</sup> Zumeist werden die beiden Ansätze unter den Namen ihrer prominenten Vertreter diskutiert. Während INMON eine zentrale Datenbank als wesentliches Merkmal des DWH ansieht (vgl. Inmon (2005): 29f.), ist diese nach KIMBALL/ROSS nicht notwendig. Vielmehr wird in seinem Ansatz ein DWH durch die Summe miteinander durch ein gleichartiges Datenschema verbundener Datenbanken (‚Data Marts‘) abgebildet (vgl. Kimball/Ross (2002): 7). Weitere Unterschiede ergeben sich aus der gewählten Modellierungsart (vgl. Abschnitt 3.1.3).

<sup>183</sup> Inmon (2005): 29. Vgl. für das Verständnis der genannten weiteren Autoren Devlin (1997): 20f.; Hansen/Neumann (2009): 1017; Mucksch/Behme (2000): 9.

- *integrated* die strukturelle und formale Vereinheitlichung von Daten, sodass im DWH lediglich ein physisches, nicht-redundantes Abbild aller Informationen aus den Quellsystemen existiert,
- *nonvolatile* eine Nicht-Veränderbarkeit der Daten, abgesehen von Fehlerfällen, die auf die Quelldaten oder den Import-Prozess zurückzuführen sind, und
- *time-variant* die Zuordenbarkeit der Daten zu Zeiträumen oder -punkten, sodass eine Analyse über die Veränderung von Daten gleicher sachlicher Bezugspunkte über die Zeit möglich ist.<sup>184</sup>

Der idealtypische Aufbau eines DWH-Systems<sup>185</sup> ist – entsprechend der Vielzahl möglicher Varianten – nicht eindeutig, zeigt jedoch typischerweise eine mehrschichtige Grundstruktur, die von der Erfassung der Daten aus den Quellsystemen über die Datenspeicherung in der zentralen Datenbank hin zur Datenbereitstellung für Analysezwecke reicht (vgl. Abb. 13).<sup>186</sup>

---

<sup>184</sup> Vgl. Inmon (2005): 29–33; Mucksch/Behme (2000): 10–13.

<sup>185</sup> Die Trennung von DWH und DWH-System wird in der Literatur nicht eindeutig vorgenommen. Im Rahmen dieser Arbeit wird die zentrale Datenbank (vornehmlich als Konzept und nicht als technische Lösung) als DWH bezeichnet und der Begriff DWH-System verwendet, wenn die (technische) Abbildung des DWH und der umliegenden Werkzeuge und Daten gemeint ist (vgl. zu der Begriffsbestimmung auch Hansen/Neumann (2009): 1019).

<sup>186</sup> Für eine alternative Referenzarchitektur, die stärker auf die Unterschiede von Daten- und Kontrollfluss eingeht, vgl. Bauer/Günzel (2013): 40–43.

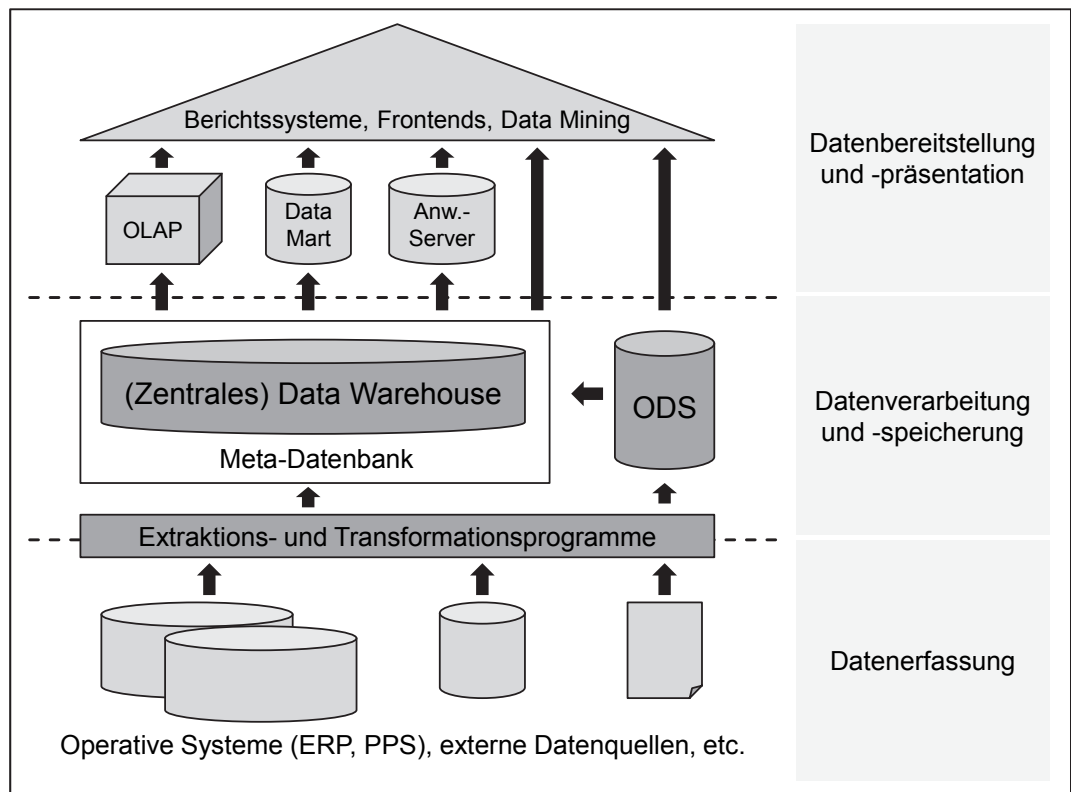


Abb. 13: Data-Warehouse-Architektur

(Quelle: Architektur in Anlehnung an Mucksch/Behme (2000): 14; Einteilung der Schichten in Anlehnung an Sinz/Ulbrich-vom-Ende (2010): 179, 182)

Die in den einzelnen Schichten durchgeführten Aufgaben können dabei wie folgt beschrieben werden:<sup>187</sup>

- Die *Datenerfassung* übernimmt alle notwendigen Schritte, um Daten verschiedener Quellsysteme (dabei kommen z. B. *Enterprise-Resource-Planning*(ERP)-Systeme, *Produktionsplanungs- und -steuerungs*(PPS)-Systeme sowie externe Daten, wie z. B. Marktdaten von Dritten, in Frage) auszulesen, in ein syntaktisch wie semantisch einheitliches Schema zu überführen und in die zentrale Datenbank zu laden. Dieser Prozess wird üblicherweise als *Extract, Transform, Load* (ETL) bezeichnet.
- Im Zuge der *Datenverarbeitung und -speicherung* werden die Daten effizient gespeichert und ggf. mit Meta-Daten angereichert. Neben der zentralen Datenbank kann ein *Operational Data Storage* (ODS) existieren. Dieses beinhaltet einen möglichst aktuellen, flüchtigen und nur in Teilen aufbereiteten Datenbestand der Quellsysteme zur direkten Weitergabe an die

<sup>187</sup> Vgl. hierzu und zum Folgenden Inmon/Strauss/Neushloss (2008): 11–13; Sinz/Ulbrich-vom-Ende (2010): 180.

analytischen Systeme, bspw. zur Verwendung in Echtzeit-Monitor-Systemen oder automatischen Analyseprozessen.

- In der *Datenbereitstellung und -präsentation* werden die gesammelten gespeicherten Daten den Anwendern zugänglich gemacht. Dabei können *Data Marts* gebildet werden, die Teilauszüge des zentralen DWH darstellen und zur Verbesserung der Zugriffsgeschwindigkeit eingerichtet werden. Neben automatisierten Analysetools und (Standard-)Berichtssystemen wird *On-Line Analytical Processing* (OLAP)<sup>188</sup> als multidimensional organisierte Zugriffstechnik auf den Datenbestand eingesetzt. *Anwendungsserver* stehen als technische Komponente zur Bereitstellung unterschiedlicher Verteilungs- und Analysefunktionen zur Verfügung.

### 3.1.3 Multidimensionale Datenhaltung und Analyse

Während in operativen Systemen zumeist ein Datensatz-orientierter Zugriff auf den Datenbestand üblich ist, herrscht im Kontext von BI-Systemen und deren Vorgängern eine multidimensional orientierte Sichtweise vor.<sup>189</sup> OLAP wird dabei als führender Begriff für eine anwenderorientierte, multidimensionale Analyse angenommen,<sup>190</sup> bietet jedoch keine einheitliche Definition eines Zielsystems. Die von CODD/CODD/SALLEY aufgestellten Grundprinzipien<sup>191</sup>, welchen Anforderungen ein OLAP-System genügen sollte – darunter breite Zugriffsmöglichkeiten, stabile Antwortzeiten und eine unbegrenzte Anzahl an Dimensionen und Aggregationsstufen der Daten –, stellen eine Grundlage für das Systemverständnis dar. Sie sind aber stark umstritten und mehrfach erweitert worden,<sup>192</sup> sodass im Folgenden die Multidimensionalität der Daten im Fokus der Betrachtung steht. Zur Veranschaulichung der multidimensionalen Datenverfügbarkeit und der darauf anwendbaren Operationen wird in der Regel ein (Hyper-)Würfel („OLAP-Cube“) gewählt, der durch beliebig viele *Dimensionen* aufgespannt wird, wobei für die graphische Repräsentation für gewöhnlich auf drei Dimensionen abgestellt wird (vgl. Abb. 14).<sup>193</sup> Die Kombinationen der *Elemente* in den Dimensionen verwei-

---

<sup>188</sup> Vgl. Abschnitt 3.1.3.

<sup>189</sup> Vgl. Gluchowski/Chamoni (2010): 199.

<sup>190</sup> Vgl. ebd.

<sup>191</sup> Vgl. Codd/Codd/Salley (1993): 12.

<sup>192</sup> Vgl. Chamoni/Gluchowski (2000): 336–343; Gluchowski/Chamoni (2010): 200–205.

<sup>193</sup> Vgl. Kemper/Baars/Mehanna (2010): 101.

sen dann auf die Datenpunkte des Würfels, die die numerischen *Fakten* (oder *Kennzahlen*) enthalten.<sup>194</sup>

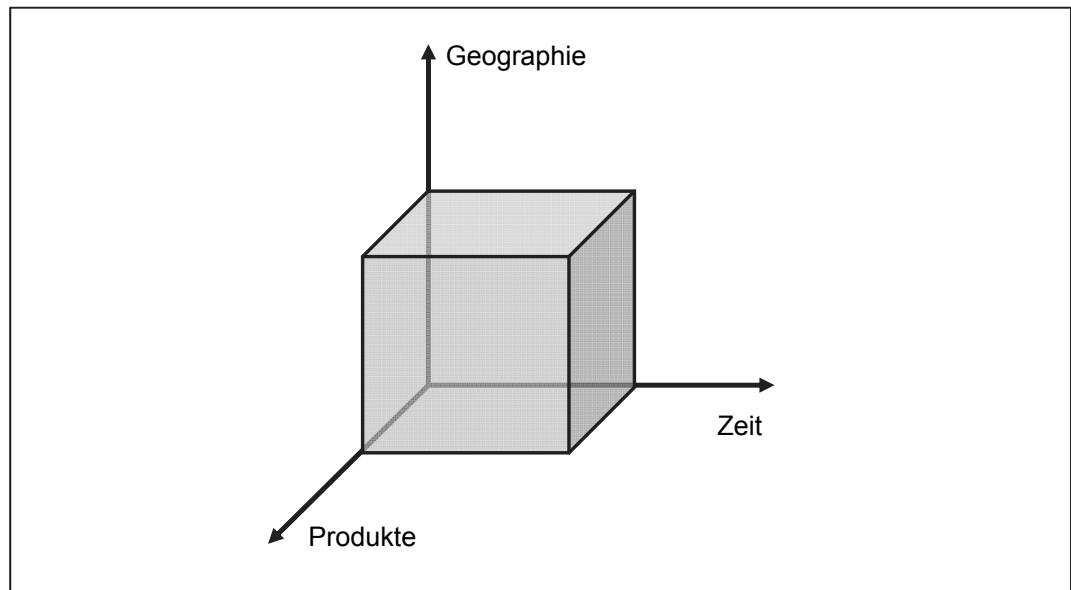


Abb. 14: Schematische Darstellung eines OLAP-Würfels  
(Quelle: In Anlehnung an Kemper/Baars/Mehanna (2010): 101)

Je nach Ausgestaltung der Dimensionen und der ihnen innewohnenden *Hierarchien* ihrer Elemente (vgl. Abb. 15), können die Navigationsmöglichkeiten in einem Würfel mehr oder weniger umfangreich genutzt werden:<sup>195</sup>

- Bei der *Pivotierung* oder *Rotation* wird der Würfel gedanklich ‚gedreht‘, sodass eine bestimmte Kombination zweier oder mehrerer Dimensionen ‚vorne‘ liegt und somit eine Detaildarstellung, z. B. in Tabellenform, anhand dieser Dimensionen erfolgt, während die Fakten über die Elemente der nicht fokussierten Dimensionen aggregiert werden.

<sup>194</sup> Die Darstellung des OLAP-Würfels zeigt deutlich die multidimensionale Sichtweise auf die Daten. Je nach gewähltem Modellierungsansatz des DWH ist diese bereits dort mehr oder weniger deutlich ausgeprägt. So wählt INMON eine Datenbank-typische normalisierte Tabellendarstellung, KIMBALL hingegen eine (in der Regel nicht-normalisierte) Darstellung in Fakten- und Dimensionstabellen. Eine hybride Variante findet sich in der Modellierung als *Data Vault*, die für Data Marts allerdings ebenfalls die bei allen Autoren (u. a. aus Performancegründen) akzeptierte Fakten-Dimensionssicht vorsieht (vgl. Krneta/Jovanovic/Marjanovic (2014): 570; sowie die originären Beschreibungen in Inmon (2005); Kimball (1996); Linstedt (2002)).

<sup>195</sup> Vgl. hierzu und zum Folgenden Kemper/Baars/Mehanna (2010): 102–105. Es existieren weitere Navigationsmöglichkeiten; die vorgestellten sind jedoch innerhalb eines ganzheitlich gestalteten Würfels die üblicherweise genannten und verwendeten (vgl. Hansen/Neumann (2009): 1026–1028).



- *Drill-down* und *Roll-up* beschreiben bei ansonsten unveränderter Dimensions- und/oder Elementauswahl die Darstellung der Fakten einer tieferen, bzw. höheren Aggregationsstufe innerhalb einer ausgewählten Dimension.
- Unter *Slicing* und *Dicing* hingegen wird das Beschränken der Auswahl auf ausgewählte Elemente/Hierarchieknoten in einer (*Slicing*) oder mehreren (*Dicing*) Dimensionen verstanden, sodass nur eine ‚Scheibe‘ oder ein ‚Teilwürfel‘ des eigentlichen Würfels dargestellt wird.

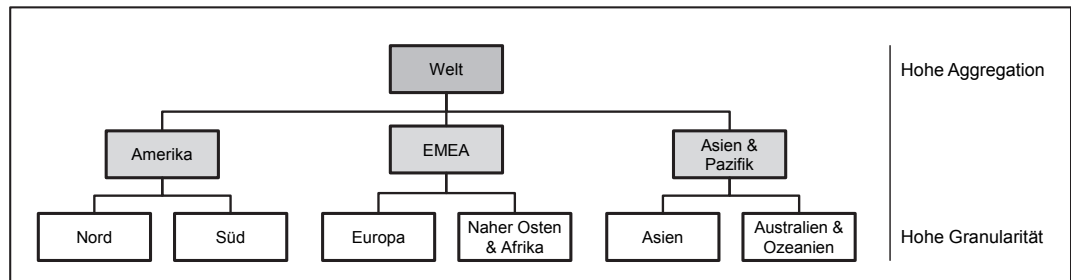


Abb. 15: Beispielhafte Darstellung einer 'Geografie'-Hierarchie

(Quelle: Eigene Darstellung)

Um eine multidimensionale Repräsentation der zu analysierenden Daten zu erreichen, ist, unabhängig von der technischen Umsetzung, ein multidimensionales Datenmodell zu entwickeln, das die Struktur der Daten und der Zugriffe vorgibt.<sup>196</sup> HAHNE unterscheidet drei Ebenen der Modellierung, die durch abnehmende Nähe zur Realwelt und zunehmende Technikbezogenheit geordnet werden können in *semantische*, *logische* und *physische* Modellierung.<sup>197</sup> Während logische und physische Modellierung bereits die zugrunde liegende Datenbanktechnologie berücksichtigen, fokussiert die semantische Modellierung das Verständnis der Daten und der Beziehungen der Realweltobjekte. Sie ist daher im Rahmen der vorliegenden Arbeit von besonderem Interesse, da nicht die (technische) Umsetzung eines DWH oder eines OLAP-Würfels im Vordergrund stehen, sondern die

<sup>196</sup> Der Modellbegriff wird in Abschnitt 4.5.1 ausführlicher erläutert. Ein Datenmodell wird hier allgemein als schematische, formale Abbildung der Realwelt verstanden, soweit sich diese auf den zu analysierenden Datenbestand bezieht (vgl. Bauer/Günzel (2013): 185–187). Die technische Umsetzung kann mit Hilfe multidimensionaler oder relationaler Datenbanken erfolgen. Die Wahl der Umsetzungsvariante ist für die in dieser Arbeit durchgeführten Überlegungen nicht von Bedeutung. Eine Einführung in die Thematik findet sich in Gluchowski/Chamoni (2010): 209–224.

<sup>197</sup> Vgl. Hahne (2010): 230f. Diese Unterteilung ist im DWH-Kontext üblich (vgl. Burmester (2011): 122–126; sowie die dort genannten Verweise).

Ausgestaltung eines mehrere BI-Systeme umfassenden Netzwerks, dessen Auswertbarkeit von der semantischen Homogenität der Daten beeinflusst wird.<sup>198</sup>

Für die semantische Modellierung stellt HAHNE mehrere Ansätze vor,<sup>199</sup> unter denen das *Application Design for Analytical Processing* (ADAPT) sowie das *Multidimensional Data Model* ( $\mathcal{MD}$ ) hervorzuheben sind. ADAPT stellt eine umfangreiche graphische Bibliothek zur Verfügung, die eine umfassende Modellierung von Dimensionen inkl. ihrer Hierarchien, Elemente und Attribute ermöglicht und auf dieser Basis auch Würfel abbilden kann (vgl. Abb. 16).<sup>200</sup>

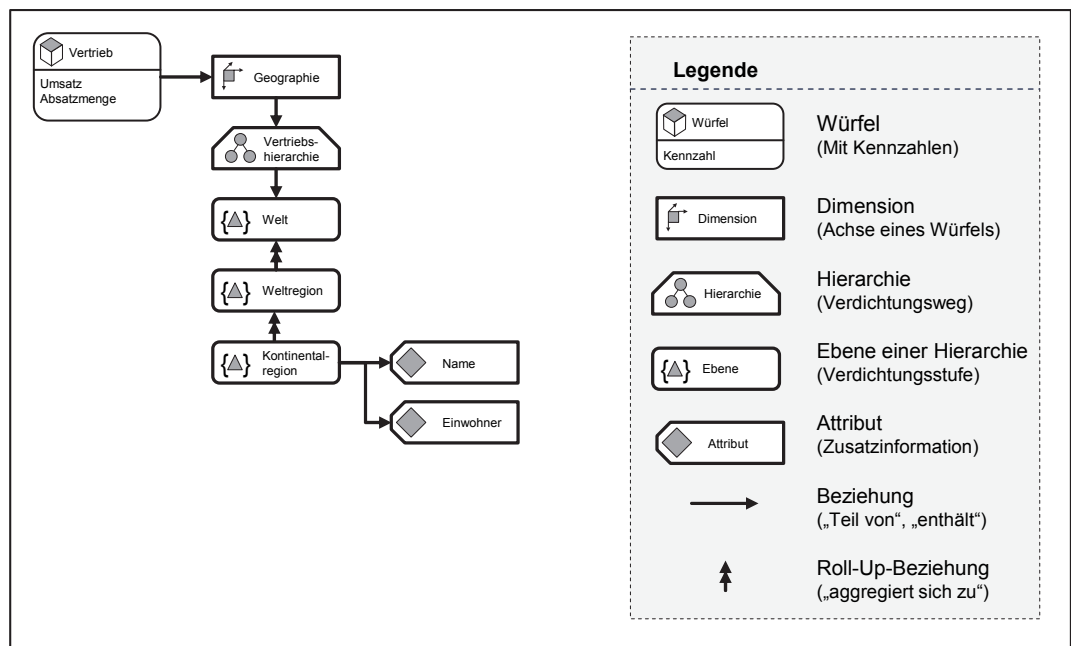


Abb. 16: Beispielhafte Modellierung einer Dimension 'Geographie' mit ADAPT als Bestandteil eines Würfels

(Quelle: Eigene Darstellung, Legende in Anlehnung an Hahne/Kurze (2012): 24)

<sup>198</sup> Vgl. hierzu insbesondere die Ausführungen in den Abschnitten 4.3 und 5.3. Die tatsächliche Implementierung und die dafür eingesetzten techniknahen Modelle sind für die Nutzbarkeit der Lösung (auch auf Netzwerkebene) von Bedeutung – z. B. hinsichtlich der erreichbaren Geschwindigkeiten – haben jedoch keine unmittelbare Auswirkung auf den Untersuchungsgegenstand dieser Arbeit (vgl. Abschnitt 1.1). Entsprechend weist TORLONE darauf hin, dass das in seiner Arbeit über die Homogenisierung von Dimensionen verwendete Modell ‚conceptual‘, d. h. konzeptionell/semantisch, und damit dem Forschungsgegenstand angemessen ist (vgl. Torlone (2008): 74). Die Grenze zur logischen Modellierung ist dennoch fließend, wie sich in dem Rückgriff auf die dem Star-Schema analoge (vgl. hierzu Hahne (2010): 243f.) Unterteilung in Dimensionen und Faktentabelle erkennen lässt, die in der ursprünglichen Formulierung des Modells besteht, wo es auch noch als ‚logical‘, d. h. logisch, gekennzeichnet wird (vgl. Cabibbo/Torlone (1998): 184).

<sup>199</sup> Vgl. Hahne (2010): 233–242. Eine noch umfangreichere, aber weniger detaillierte Aufstellung von Modellierungsmethoden findet sich bei PEDERSEN (vgl. Pedersen (2013): 13–17).

<sup>200</sup> ADAPT wird in dieser Arbeit für die graphische Darstellung der verwendeten Modelle genutzt (vgl. Abschnitt 6.2).

Das  $\mathcal{MD}$  liefert eine weniger mächtige graphische Repräsentation, die nicht so sehr an fachlichen Bedarfen orientiert ist,<sup>201</sup> ist aber durch eine klare formale Definition gekennzeichnet, die das Verständnis von Dimensionen, Hierarchien und Funktionen auf diesen expliziert und die Formulierung einer Dimensionsalgebra ermöglicht:<sup>202</sup>

**Definition (1)** Ein *Dimensionsschema*  $S(d)$  besteht aus

- einer Menge an Hierarchieebenen  $L = \{l_1, \dots, l_n\}$  und
- einer Ordnung  $\preceq$  auf  $L$ , mit der Bedeutung, dass für  $l_1 \preceq l_2$  gilt: Elemente in  $l_1$  werden zu Elementen in  $l_2$  aggregiert, d. h.  $\preceq$  ermöglicht ein ‚Roll-Up‘ von  $l_1$  zu  $l_2$ .<sup>203</sup>

Es wird vorausgesetzt, dass  $L$  eine endliche Menge ist und ein Grundelement  $\perp$  (in Bezug auf  $\preceq$ ) enthält.

**Definition (2)** Eine *Dimensionsinstanz*  $I(d)$  besteht aus

- einer Funktion  $m$ , die Realweltelemente mit Hierarchieebenen assoziiert und
- einer Gruppe von Funktionen  $\rho$ , die eine Roll-Up-Funktion  $\rho^{l_1 \rightarrow l_2}: m(l_1) \rightarrow m(l_2)$  enthält für alle Paare  $l_1 \preceq l_2$ .

**Definition (3)** Eine *Dimension*  $d$  besteht aus einem Dimensionsschema  $S(d)$  und einer Dimensionsinstanz  $I(d)$ .

**Definition (4)** Ein *Würfel*  $c$  besteht aus

- einer Menge von Dimensionen  $D = \{d_1, \dots, d_n\}$ ,

<sup>201</sup> Vgl. Hahne (2010): 238, 255.

<sup>202</sup> Vgl. hierzu und zum Folgenden Cabibbo/Torlone (1998): 186–188; Torlone (2008): 73–76. Aufgeführt sind lediglich die Definitionen, die dem Verständnis von  $\mathcal{MD}$  zuträglich sind und/oder für die formale Bestimmung von Dimensionsähnlichkeiten/-überführungen sowie Gütekriterien notwendig sind (vgl. insbesondere Abschnitt 4.3.3). Bei Differenzen in der Formulierung zwischen den genannten Publikationen wird Torlone (2008) als führend angenommen. Die Notation wird mittlerweile in ihrer Grundform als für OLAP-bezogene Datenmodelle üblich angesehen (vgl. Vaisman/Minuto Espil/Paradela (2009): 234). Ergänzend sei darauf hingewiesen, dass  $\mathcal{MD}$  hier explizit von der Bezeichnung MD als Abkürzung für das ‚multidimensional modeling‘ abgegrenzt wird.

<sup>203</sup> Die Begriffe ‚aggregiert‘ und ‚verdichtet‘ werden im gegebenen Kontext synonym verwendet. Für eine Einführung in die Mengenlehre und Ordnungsrelationen vgl. Auer/Seitz (2013): 9–14 und Berghammer (2012): 4–10.

- einem Faktenschema  
 $f[A_1:l_1, \dots, A_n:l_n] \rightarrow \langle M_1:\tau_1, \dots, M_m:\tau_m \rangle$  für das gilt,  
 dass jedes  $A_i$  ein eindeutig unterscheidbarer Attributname,<sup>204</sup> jedes  $l_i$  eine Hierarchieebene einer Dimension in  $D$ ,  
 jedes  $M_j$  eine Kennzahl und jedes  $\tau_j$  die Beschreibung eines numerischen Basisdatentyps ist, sowie
- einer Instanz als Funktion, die jeder **Koordinate** einen **Fakt** zuweist. Dabei ist eine Koordinate definiert als Tupel über alle  $A_i$  in  $f$ , das für jedes  $A_i$  ein Element von  $l_i$  bestimmt. Ein Fakt ist definiert als Tupel über alle  $M_j$  in  $f$ , das für jedes  $M_j$  einen Wert des Typs  $\tau_j$  bestimmt.

Exemplarisch lässt sich auf Basis von Abb. 15 und Abb. 16 ein Würfel ‚Vertrieb‘ beschreiben, dessen einzige Dimension ‚Geografie‘ die geordnete Menge der Hierarchieebenen ‚Kontinentalregion‘  $\leq$  ‚Weltregion‘  $\leq$  ‚Welt‘ aufweist, wobei ‚Kontinentalregion‘ das Grundelement  $\perp$  darstellt. Die Instanz des Würfels legt die Zuweisung von Koordinaten, d. h. Elementkombinationen (hier durch  $|D| = 1$  gleichbedeutend mit Elementen, z. B. „Europa“), zu dem Fakt ‚Umsatz‘ und ‚Absatzmenge‘ fest.<sup>205</sup>

Der in Definition (4) dargestellte Würfel lässt sich in einem DWH-System sowohl in multidimensionaler Form als auch in relationaler Form materialisieren, wobei auch Zwischenvarianten möglich sind. Je nach Speicherart wird dies dann im Kontext des OLAP als *relationales OLAP* (ROLAP), *multidimensionales OLAP* (MOLAP) oder *hybrides OLAP* (HOLAP) bezeichnet.<sup>206,207</sup> Das *Star-Schema*

<sup>204</sup> Die Annahme eindeutig unterscheidbarer Attributnamen wird von den Autoren des Originals getroffen. Für die hier wiedergegebene formale Definition ist diese Einschränkung vonnöten. In der Praxis können bei der Instanziierung des Modells ggf. identische Attributnamen im Sinne von Elementnamen in verschiedenen Dimensionen auftreten, z. B. bei numerischen Schlüsseln (bspw. bei Kundennummer und Postleitzahl). Dies kann umgangen werden durch die Hinzunahme des Dimensionsnamens in die Elementbezeichnung, sodass die Forderung nach einem eindeutigen Attributnamen die Nutzbarkeit des Modells nicht beeinflusst und keine Beschränkung der theoretischen Überlegungen darstellt.

<sup>205</sup> Sprachlich korrekt ist hier der Begriff ‚Fakt‘ anstelle des Plurals, da der ‚Fakt‘ das Tupel der Kennzahlen darstellt. Diese Unterscheidung wird nicht einheitlich getroffen, sodass auch einzelne Kennzahlen als Fakten bezeichnet werden (vgl. Hahne (2010): 237).

<sup>206</sup> Vgl. Kemper/Baars/Mehanna (2010): 106f.

stellt dabei eine für DWH-Systeme typische Form der relationalen Speicherung dar: In einer zentralen Faktentabelle werden die eigentlichen Kennzahlen (z. B. ‚Anzahl Verkäufe‘ oder ‚Umsatz‘ eines Artikels) gespeichert und mit einem Primärschlüssel, d. h. einer eindeutig identifizierenden Attributkombination, versehen, dessen Einzelattribute sich wiederum aus den Datensätzen der beschreibenden Dimensionstabellen ableiten lassen.<sup>208</sup> Abb. 17 zeigt eine beispielhafte Darstellung eines Würfels mit drei Dimensionen, die zwei Fakten identifizieren.

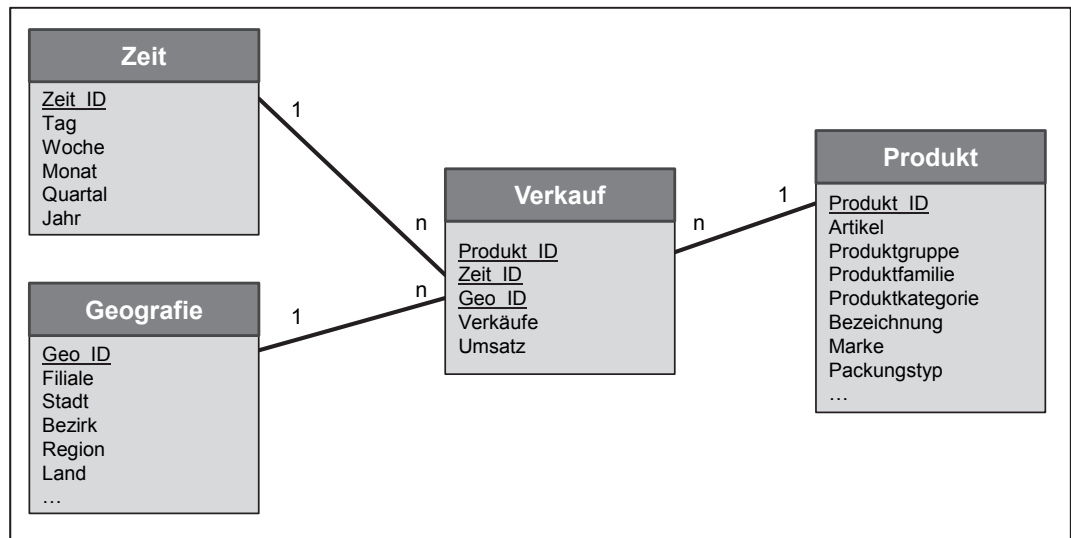


Abb. 17: Beispiel eines Star-Schemas<sup>209</sup>

(Quelle: Bauer/Günzel (2013): 245)

Neben den durch menschliche Auswahl und Interpretation geprägten Analysen im OLAP-Kontext können auf den Daten im DWH auch maschinelle Analysen durchgeführt werden. Dies geschieht insbesondere im Kontext des Knowledge-Discovery-in-Databases(KDD)-Prozesses (vgl. Abb. 18).

<sup>207</sup> Die Implementierung ist für die semantische Modellierung nicht erheblich, hat jedoch erhebliche Auswirkungen auf die einsetzbaren Techniken zur Datenabfrage und -analyse. Die Umsetzung in relationalen Tabellen ermöglicht bspw. den Zugriff auch auf relationaler Ebene und damit den Rückgriff auf erprobte Techniken relationaler Datenbankmanagementsysteme, wie es im Rahmen des externen Zugriffs im Bereich verteilter Data-Warehouse-Systeme und -Netzwerke verwendet wird. Der Einsatz von ROLAP ist daher teilweise für weitergehende Konzepte Voraussetzung (vgl. Abschnitt 4.2).

<sup>208</sup> Vgl. Bauer/Günzel (2013): 243–246.

<sup>209</sup> Die Ordnungszahlen (1, n) an den Beziehungen geben an, dass zu jedem Datensatz der Faktentabelle genau 1 Datensatz in jeder Dimensionstabelle existieren muss und dass jeder Datensatz der Dimensionstabelle von null bis beliebig vielen Datensätzen der Faktentabelle referenziert wird. Zu der als Grundlage für diese Modellierung verwendete ‚Entity-Relationship-Modellierung‘ vgl. Chen (1976).

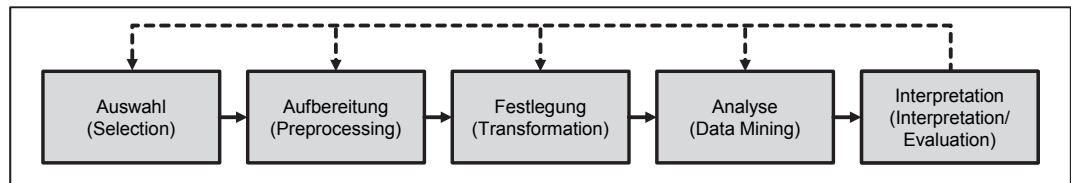


Abb. 18: Prozess des Knowledge Discovery in Databases

(Quelle: In Anlehnung an Fayyad/Piatetsky-Shapiro/Smyth (1996b): 41)

KDD kann nach FAYYAD/PIATETSKY-SHAPIRO/SMYTH dabei verstanden werden als „non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data“<sup>210</sup>. Die Autoren grenzen Data Mining dabei explizit als Teilschritt ab und bezeichnen es als die Anwendung spezifischer Algorithmen zur Mustererkennung.<sup>211</sup> Die so beschriebenen Verfahren dienen gleichermaßen einer Deskription der bestehenden Daten wie einer Prädiktion zukünftig anfallender.<sup>212</sup> Dabei werden in der Praxis typischerweise Verfahren zur Segmentation, Klassifikation, Vorhersage und Assoziation eingesetzt.<sup>213</sup> Im Rahmen der Entscheidungsunterstützung können so anwendungsfallgetriebene Maßnahmen eingeleitet werden um auf bestimmte Faktenkonstellationen hin- oder ihnen entgegenzuwirken.<sup>214</sup>

### 3.2 Gruppenentscheidungsunterstützung

Während der vorherige Abschnitt Individualentscheidungen und die dafür mögliche Unterstützung thematisiert, werden im Folgenden kooperative Entscheidungen im Sinne von Gruppenentscheidungen und den unterstützenden Systemen betrachtet. Die Basisannahme für die Rechtfertigung von Gruppenentscheidungen ist dabei, dass eine Menge von Individuen in Summe zu einer ‚besseren‘ Entscheidung kommen kann als der Einzelne.<sup>215</sup> Im Folgenden werden daher zunächst Gruppen und Gruppenentscheidungen thematisiert, dann die Möglichkeiten zur technischen Unterstützung skizziert, wie sie seit mehreren Jahrzehnten unter

<sup>210</sup> Fayyad/Piatetsky-Shapiro/Smyth (1996a): 6.

<sup>211</sup> Vgl. Fayyad/Piatetsky-Shapiro/Smyth (1996b): 39.

<sup>212</sup> Vgl. Fayyad/Piatetsky-Shapiro/Smyth (1996a): 12.

<sup>213</sup> Vgl. Bankhofer/Vogel (2008): 255.

<sup>214</sup> Typische Beispiele liegen in der Kundensegmentierung im Marketing oder in der Klassifizierung von Risiken, z. B. in der Versicherungsbranche (vgl. Chamoni/Beeckmann/Bley (2010): 331f.) LIU/DANIELS weisen darauf hin, dass dies auch ein geeignetes Anwendungsgebiet für unternehmensübergreifende Datenanalysen darstellt (vgl. Liu/Daniels (2012): 1).

<sup>215</sup> Vgl. Surowiecki (2007): 11.

dem Begriff der Group Decision Support Systems (GDSS) geführt werden und im Anschluss die Implikationen technischer Entwicklungen im Bereich der ‚sozialen Medien‘ diskutiert.

### 3.2.1 Gruppenentscheidungsprozesse

Die Betrachtung von Gruppenentscheidungen und -entscheidungsprozessen ist unabhängig von der Durchführung und systemischen Unterstützung zunächst ein Forschungsfeld der Sozialpsychologie. Um die Anforderungen an Systeme zur Gruppenentscheidungsunterstützung zu verdeutlichen, wird im Folgenden eine kurze Einführung in die Thematik gegeben. Auf eine ausführliche Betrachtung der theoretischen Grundlagen zur (sozial-)psychologischen (Gruppen-)Entscheidungsforschung sowie eine Einführung in den historischen Hintergrund und die Entwicklung des Feldes wird dabei verzichtet.<sup>216</sup>

Der Begriff der *Gruppe* kann unterschiedlich verstanden werden und setzt je nach Ansatz lediglich eine gemeinsame Wahrnehmung einer Gemeinschaft oder eine gleichartige Motivation und Zielsetzung voraus.<sup>217</sup> Diese Arbeit folgt auf Grund der Ausrichtung an Entscheidungsprozessen der Definition von DESANTIS/GALLUPE und versteht als entscheidungsorientierte Gruppe zwei oder mehr Personen, die gemeinsam verantwortlich für Problemerkennung, Lösungsgenerierung und -evaluation sowie die Beschreibung einer Strategie zur Lösungsimplementierung sind. Physische Nähe wird dabei nicht vorausgesetzt, wohl aber Kenntnis voneinander sowie ein gemeinschaftliches Zugehörigkeitsempfinden zu den Entscheidungsträgern.<sup>218</sup>

Die beschriebenen Aufgaben entsprechend damit weitgehend den Phasen im Grundmodell eines Entscheidungsprozesses, wie er auch bei Individualentscheidungen zum Tragen kommt (vgl. Abb. 19). Von diesem *Gruppenentscheidungsprozess* ist der *kollektive Entscheidungsprozess* abzugrenzend, da er zwar berücksichtigt, dass mehrere Personen (oder Organisationsteile) an der endgültigen Entscheidung beteiligt sind, jedoch keine Wahrnehmung der Beteiligten als Gemeinschaft vorliegen muss.<sup>219</sup>

---

<sup>216</sup> Für vertiefende Informationen und einen Literaturüberblick vgl. Beck (2001): 32–37.

<sup>217</sup> Vgl. Ellis/Fisher (1994): 4f.

<sup>218</sup> Vgl. DeSanctis/Gallupe (1987): 590.

<sup>219</sup> Vgl. Kirsch (1971): 53f.

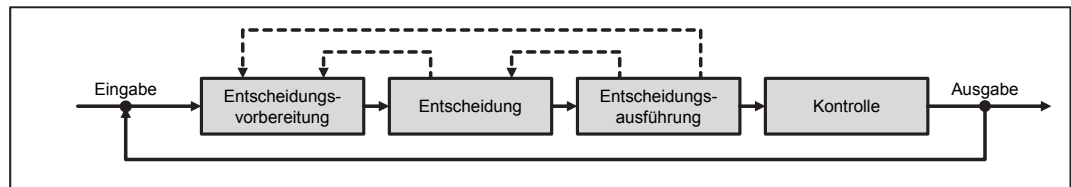


Abb. 19: Grundmodell des Entscheidungsprozesses

(Quelle: Pfohl/Braun (1981): 103)

PFOHL/BRAUN weisen darauf hin, dass in Entscheidungsprozessen mit mehreren Beteiligten – z. B. durch den Einfluss von Machtprozessen – häufig Rücksprünge in frühere Phasen des Prozesses stattfinden.<sup>220</sup> Diese u. a. durch Anpassungen von Individualmeinungen verursachten Änderungen im Gruppengefüge müssen im Prozess berücksichtigt werden.<sup>221</sup>

Verbesserungen, z. B. durch systemische Unterstützung, sollen die *Qualität*<sup>222</sup> der Gruppenentscheidung erhöhen. Zur Messung dieser Qualität schlägt RUBIN sechs Dimensionen vor:<sup>223</sup>

- Die *Effizienz* einer Gruppe steigt, wenn der Zeitraum bis zur Erreichung einer gemeinsamen Entscheidung kleiner wird.
- Eine *gründliche Entwicklung und Analyse von Alternativen* wird durch die Gruppe und ihre Mitglieder sichergestellt.
- Die *Fairness*<sup>224</sup> einer Entscheidung ist hoch, wenn die Interessen aller Gruppenmitglieder und die Interessen aller von der Entscheidung beeinflussten Nicht-Gruppenmitglieder berücksichtigt werden und die Entscheidung beeinflussen können.
- Die *Zufriedenheit und Moral der Gruppenmitglieder* steigt, wenn jedes Mitglied sich eingebunden und wertgeschätzt fühlt. Der Einfluss auf die Qualität der Entscheidung ist indirekt. Zum einen wird eine höhere Akzeptanz geschaffen, zum anderen wird die Fähigkeit der Gruppe in Bezug auf

<sup>220</sup> Vgl. Pfohl/Braun (1981): 118.

<sup>221</sup> Vgl. Vetschera (1991): 5.

<sup>222</sup> Zur genauen Begriffsbestimmung von ‚Qualität‘ und den damit verbundenen Schwierigkeiten vgl. Abschnitt 5.1.

<sup>223</sup> Vgl. hierzu und zum Folgenden Rubin (1984): 18–21.

<sup>224</sup> Der Begriff der ‚Fairness‘ kann als ‚anständiges Verhalten‘ übersetzt werden, hat mittlerweile jedoch auch Einzug in den deutschen Sprachgebrauch gehalten und wird in der Regel ohne Übersetzung verwendet (vgl. Duden (2015a)).



die nächste zu treffende Entscheidung durch die Bereitschaft zur erneuten Partizipation erhöht.

- Die *Effektivität der Führung* wird durch die Fähigkeit der Gruppe, gute Entscheidungen zu treffen, erhöht, da bei gleichem Aufwand des Moderators/der Führung eine bessere durch ihn/sie zu vertretende Entscheidung erreicht wird.
- Der *Lerneffekt der Gruppe über die Zeit* ist ein ebenfalls indirektes Maß, das (ähnlich der Zufriedenheit) bei repetitiven Entscheidungsprozessen zu einer schnellen Verbesserung nachfolgender Entscheidungen führt.

Erkennbar ist der Unterschied direkter und indirekter Auswirkungen auf die Entscheidung, die RUBIN immer vor dem Hintergrund mehrfach durchgeführter Entscheidungsprozesse betrachtet. Die indirekten Effekte sind im Falle einfach durchgeführter Entscheidungen entsprechend zu vernachlässigen. Allgemein ist das Erreichen eines guten Wertes in der einen Dimension ggf. mit Absinken des Wertes in einer anderen Dimension verbunden.<sup>225</sup> Deutlich wird, dass die vorgestellten Qualitätsaspekte nicht vollständig quantifizierbar sind und keine direkte Umsetzung in IT-Systemkennzahlen zulassen. Sie bieten dennoch einen Kontext, der z. B. in den Überlegungen zu Vertrauen und Verlässlichkeit Eingang in die Forschung zu unternehmensinternen und unternehmensübergreifenden BI-Systemen findet.<sup>226</sup>

### 3.2.2 Systemische Unterstützung

VETSCHERA unterscheidet drei Grundgedanken, die vor dem Hintergrund von existierenden Individualentscheidungsunterstützungssystemen und Gruppenentscheidungsprozessen zu der Entwicklung von GDSS geführt haben.<sup>227</sup>

- Gruppenentscheidungsprozesse können im Rahmen der *Spieltheorie* behandelt werden. Dabei handelt es sich um die Untersuchung von Entscheidungsprozessen mehrerer Individuen, die aufeinandertreffen, wobei die

---

<sup>225</sup> Vgl. Rubin (1984): 19. Ein ähnlicher Effekt wird durch das ‚magische Dreieck‘ beschrieben, das zumeist im Projektmanagement eingesetzt wird und den Konflikt der Erreichung einer kurzen Zeit, einer hohen Qualität oder Leistung und geringer Kosten/Ressourcenbedarfe der Umsetzung des Projektauftrags beschreibt (vgl. Keßler/Winkelhofer (2004): 55; Stein (2009): 44).

<sup>226</sup> Vgl. Liu/Daniels (2012): 3.

<sup>227</sup> Vgl. hierzu und zum Folgenden Vetschera (1991): 1–4.

*kooperative Spieltheorie* miteinbezieht, dass es zu Abhängigkeiten zwischen den Beteiligten kommen kann.<sup>228</sup>

- *Multikriterielle Zielsetzungen* müssen bei Entscheidungen berücksichtigt werden, die verschiedene Umweltaspekte gleichzeitig beeinflussen. Dies kann einzelne Entscheider betreffen, die einen Kompromiss bei ihrer Entscheidung finden müssen, trifft jedoch genauso auf Gruppenentscheidungen zu, in denen die Gesamtheit der Beteiligten die möglicherweise unterschiedlichen Präferenzen und Lösungsansätze zu einem Ergebnis vereinen muss.
- *Decision Support Systems* bieten eine systemische Grundlage und fokussieren stärker als die entscheidungstheoretischen Konzepte auf eine Unterstützung des Anwenders bei der Entscheidungsfindung als auf eine strukturierte Ergebnishinführung im Sinne eines Ergebnis-festlegenden Modells. Entsprechend orientieren sie sich stärker am Entscheidungsprozess als an der ‚richtigen‘ Entscheidung.<sup>229</sup> Nach GRAY entspricht ein GDSS genau dann einem DSS, wenn die Gruppengröße der Nutzer eins beträgt.<sup>230</sup>

Auf Basis der Erfahrungen mit DSS und der sozialwissenschaftlich orientierten Erkenntnisse zu Gruppenentscheidungsprozessen<sup>231</sup> haben DESANCTIS/GALLUPE für die Forschung zu GDSS ein mehrdimensionales Gerüst geschaffen, das die Aspekte des Feldes beschreibt (vgl. Abb. 20).

---

<sup>228</sup> Vgl. Wessler (2012): 5f; sowie die Ausführungen zur Kooperationsmotivation in Abschnitt 2.1.

<sup>229</sup> Vgl. Vetschera (1991): 4.

<sup>230</sup> Vgl. Gray (1987): 237.

<sup>231</sup> Vgl. Abschnitt 3.2.1.

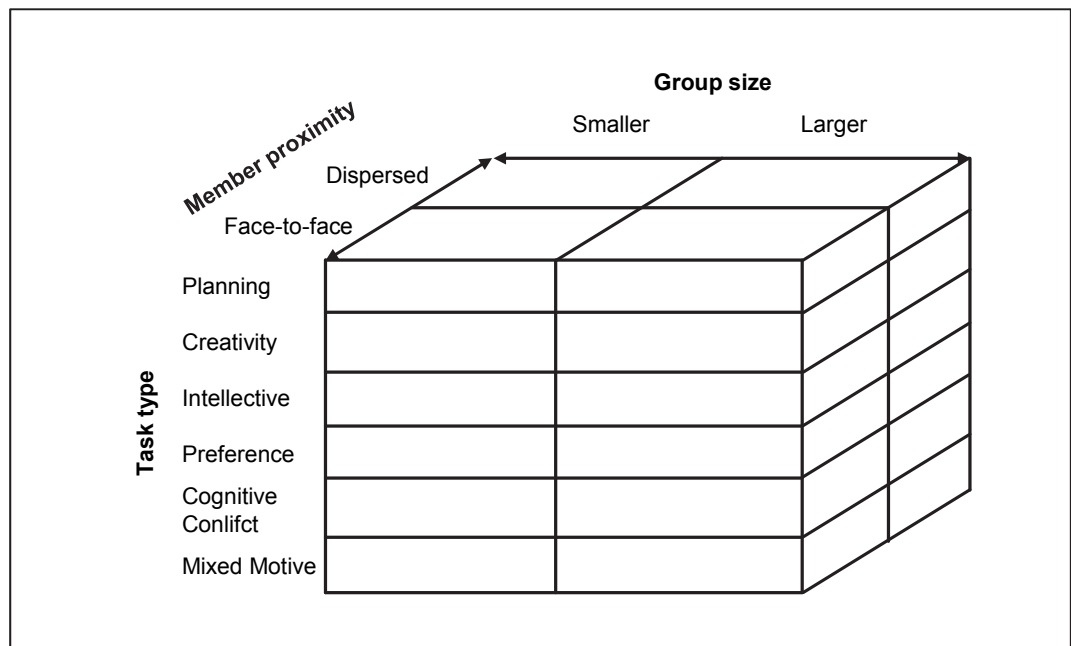


Abb. 20: Dimensionen der Forschung zu Gruppenentscheidungsunterstützungssystemen

(Quelle: DeSanctis/Gallupe (1987): 591)

Während die Aufgabenarten dabei an typische Aufgaben aus der betrieblichen Praxis angelehnt sind, unterscheiden die Autoren bei der Gruppengröße und der (räumlichen) Nähe der Gruppenmitglieder jeweils zwei deutlich umschriebene und entgegengesetzte Ausprägungen. Die Unterscheidung zwischen kleinen und großen Gruppen ist dabei ein mehrfach auftretender Aspekt in der wissenschaftlichen Literatur.<sup>232</sup> HOWARD weist jedoch darauf hin, dass gerade die Entwicklung von GDSS und die systemische Unterstützung von Gruppenentscheidungsprozessen mit ihrer Interaktionsstärkung der Beteiligten diesen Aspekt in den Hintergrund drängen.<sup>233</sup>

Die Unterscheidung von räumlich geeinter und räumlich getrennter Arbeit treffen NUNAMAKER ET AL. ebenfalls, ergänzen diese jedoch noch um die Möglichkeit einer zeitlichen Entkopplung (vgl. Abb. 21), wie sie auch im Kontext von Groupware, d. h. systemischer Unterstützung reinen Zusammenarbeitens,<sup>234</sup> diskutiert wird.<sup>235</sup>

<sup>232</sup> Vgl. Franz (2012): 4f.; Rubin (1984): 54f.; Abschnitt 3.2.1.

<sup>233</sup> Vgl. Howard (1994): 20f. Als ‚kleine‘ Gruppen bezeichnet die Autorin Gruppen von bis zu vier Teilnehmern, als ‚große‘ Gruppen solche ab fünf bis sieben Personen.

<sup>234</sup> Vgl. Matt (2012): 164.

<sup>235</sup> Vgl. Fischer et al. (2008): 179f.

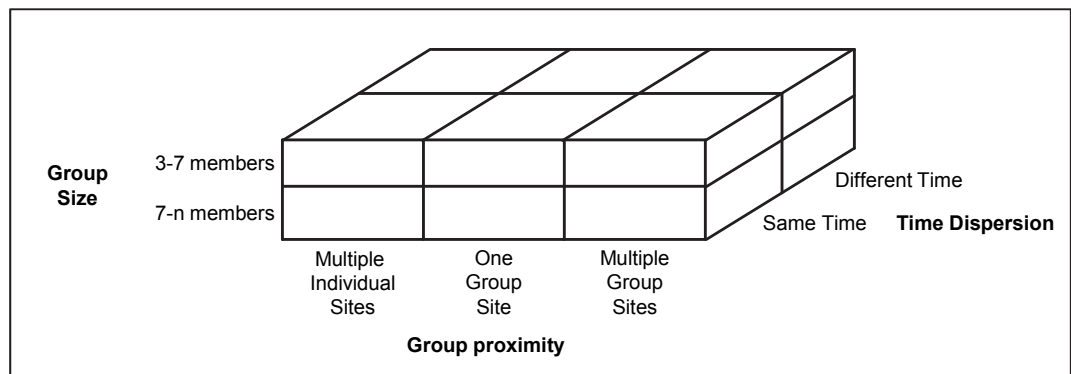


Abb. 21: Klassifikationen von Systemen zur Gruppenzusammenarbeit

(Quelle: In Anlehnung an Nunamaker et al. (1991): 44)

Vor dem Hintergrund der durchzuführenden Aufgaben, insbesondere im Rahmen der Lösungsfindung und -bewertung,<sup>236</sup> können GDSS in drei Ansätze unterschieden werden, die von POWER alle als Communication-driven DSS verstanden werden:<sup>237</sup>

- *Level-1-GDSS* stellen technische Möglichkeiten zur Barrierenreduzierung im Kommunikationsprozess zur Verfügung. Ziel ist es, den Informationsaustausch zu erleichtern und eine gemeinsame Entscheidungsfindung dadurch zu begünstigen, dass der Kenntnisstand homogen ist.
- *Level-2-GDSS* stellen Modelle und Methoden, wie sie aus DSS bekannt sind, bereit. Dazu gehören z. B. Planungsmodelle und Risikoanalysen. Weiterhin können sie eine vorgegebene Struktur eines Entscheidungsprozesses abbilden und den Entscheidungsprozess der Gruppe damit gezielt führen, so wie es normative Ansätze im Group Decision Making vornehmen.
- *Level-3-GDSS* stärken die Rolle des Systems im Entscheidungsprozess und kontrollieren z. B. die Kommunikationswege und -zeiten der Beteiligten.

<sup>236</sup> Vgl. Abschnitt 3.2.1

<sup>237</sup> Vgl. zur Unterscheidung der Ansätze und zur folgenden Aufstellung DeSanctis/Gallupe (1987): 593f. Die Arbeit von DeSanctis/Gallupe kann als führendes und meistzitiertes Werk zu dem Thema gelten (vgl. Markus/Silver (2008): 611; sowie die Anzahl von Zitierungen durch Dritte (Google Scholar, Suche nach „Group Decision Support Systems“, 1997 ausgewiesene Zitierungen der Arbeit. Die nächsthäufig zitierten Beiträge weisen 727, bzw. 498 Zitierungen auf. Stand: 05.12.2013)). Zur Einordnung der DSS-Typen vgl. Power (2009); sowie Abschnitt 3.1.1.

Der Effekt von GDSS (Level 2) wurde empirisch beobachtet und zeigt, dass die in Abschnitt 3.2.1 diskutierten Qualitätsmaße unterschiedlich beeinflusst werden. So werden die Effizienz der Entscheidung, die Zufriedenheit der Gruppenmitglieder mit Prozess und Ergebnis sowie die Organisation des Prozesses erhöht, es ist jedoch auch ein Nachlassen der Gründlichkeit in der Ausarbeitung der Entscheidungen zu beobachten.<sup>238,239</sup>

### 3.2.3 Veränderungen durch den Einfluss sozialer Medien

Die Entwicklung sozialer Medien wird in der Regel unter dem englischsprachigen Begriff *Social Media* betrachtet. Er beschreibt technische Einrichtungen, die es Individuen erlauben, bestimmte Tätigkeiten unabhängig von räumlicher Nähe gemeinschaftlich durchzuführen, wobei zumeist das Internet und insbesondere die Technologien des Web 2.0 als Plattform angenommen werden.<sup>240</sup> *Web 2.0* ist der in diesem Kontext gebräuchliche Begriff für die Verwendung des Internets als interaktives Medium zwischen ursprünglichem Inhaltsanbieter und Inhaltsnutzer, sodass der eigentliche Konsument durch das Beisteuern von eigenen Inhalten eine aktive Rolle einnimmt.<sup>241</sup> Wird die technologische Basis nicht auf das Internet eingeschränkt, findet der Begriff *Social Software* Verwendung.<sup>242</sup> Im Unternehmenskontext wird auch der Begriff *Enterprise 2.0* verwendet, bei dem McAfee explizit darauf hinweist, dass mit den verwendeten Methoden nicht nur die unternehmensinterne, sondern auch die unternehmensübergreifende Zusammenarbeit gestärkt werden kann.<sup>243</sup>

Als eine der Hauptsystemklassen von Social Media kann die der *Online Social Networks* (OSN) betrachtet werden, da sie sowohl im privaten als auch berufli-

---

<sup>238</sup> Vgl. Poole et al. (1993): 176.

<sup>239</sup> Diese Arbeit fokussiert im Kontext der DSS auf den Bereich BI. Die Ausführungen zu GDSS sind entsprechend auf die in Abschnitt 3.1 beschriebenen BI-Systeme zu beziehen. Vor diesem Hintergrund werden die Auswirkungen der technologischen Entwicklung sozialer Medien auf BI-Systeme und die damit verbundene Unterstützung von kooperativer Arbeit gezeigt. Die Verwendung von BI-Modellen und -Methoden kombiniert mit kommunikationsfördernden Werkzeugen weist dabei die größte Ähnlichkeit zu Level-2-GDSS auf, sodass diese im Folgenden stellvertretend für GDSS gelten.

<sup>240</sup> Vgl. Heßler/Mosebach (2013): 279; Kaplan/Haenlein (2010): 61. In Teilen wird Social Media als über das technologische Verständnis hinausgehend verstanden und bezieht die inhaltliche und gestalterische Perspektive mit ein (vgl. Hettler (2010): 14).

<sup>241</sup> Vgl. O'Reilly (2005); Kaplan/Haenlein (2010): 60f. Der so erzeugte Inhalt wird auch als *User Generated Content* bezeichnet (vgl. ebd.).

<sup>242</sup> Vgl. Stegbauer/Jäckel (2008): 7.

<sup>243</sup> Vgl. McAfee (2006).

chen Umfeld häufig zum Einsatz kommt.<sup>244</sup> Ein OSN ist dabei ein Web-Service, der es den Nutzern erlaubt, ein Profil von sich zu erstellen, sich mit anderen Nutzern explizit zu verbinden und die Verbindungen auch über mehrere Ebenen nachzuvollziehen.<sup>245</sup>

Die in Social Media allgemein und in OSN im Besonderen verfügbaren Technologien entsprechen in ihrer Nutzung für die Unterstützung gemeinschaftlicher Arbeit in Teilen den für GDSS vorgesehenen Mitteln.<sup>246</sup> Sie werden aber auch zunehmend im Rahmen der BI eingesetzt, um diese datengetriebene Form der DSS hinsichtlich gemeinschaftlicher Entscheidungsfindung zu unterstützen. So beschreiben DAYAL ET AL. eine Lösung zum Einsatz von BI-Instrumenten im Rahmen kooperativer Systeme.<sup>247</sup> Auch die reine Ergänzung von existierenden BI-Systemen durch OSN und die in ihnen bestehenden Personennetzwerke zur Diskussion und Entscheidungsfindung in unternehmensinternen oder -übergreifenden Gruppen wird diskutiert.<sup>248</sup> Für diese Ansätze verwenden die Autoren den Begriff Collaborative Business Intelligence, der im folgenden Abschnitt ausführlich diskutiert wird.

### 3.3 Collaborative Business Intelligence

Die beschriebenen Anwendungen zur Entscheidungsunterstützung in Unternehmen geraten zunehmend unter den Druck einer Öffnung nach außen, um die gemeinschaftliche Problemlösung zu unterstützen und z. B. Wertschöpfungsketten einen analytischen Mehrwert zu verschaffen.<sup>249</sup> Im Zusammenhang mit Planungs- und Steuerungsprozessen wird dies bereits durch Ansätze wie CPFR realisiert.<sup>250</sup>

Neben der Forschung hat auch die Praxis diese Problematik adressiert. Das Marktforschungsunternehmen GARTNER<sup>251</sup> veröffentlicht jährlich für das Fachgebiet BI (und andere Gebiete aus dem Technologie-Bereich) Berichte über aktuelle Entwicklungen und Trends. Bestandteil dieser Veröffentlichungen ist der Hype Cycle (oder Hype-Zyklus) des jeweiligen Gebietes. In den Berichten von 2011 bis 2013

---

<sup>244</sup> Vgl. Busemann/Gscheidle (2012): 391; LAB & Company (2012).

<sup>245</sup> Vgl. Boyd/Ellison (2007): 211; Mislove et al. (2007): 30.

<sup>246</sup> Vgl. Abschnitt 3.2.2.

<sup>247</sup> Vgl. Dayal et al. (2008): 24.

<sup>248</sup> Vgl. Kaufmann/Chamoni (2013): 127.

<sup>249</sup> Chamoni/Gluchowski (2010): 15.

<sup>250</sup> Vgl. Abschnitt 2.3.1.

<sup>251</sup> <http://www.gartner.com>.

wird Collaborative Decision Making (CDM) als Technologie im Stadium ‚Technology Trigger‘ geführt,<sup>252</sup> in dem die Technologie steigende Aufmerksamkeit erfährt, aber noch unausgereift sein kann.<sup>253</sup> In diesem Zusammenhang wird auch der Begriff Collaborative Business Intelligence aufgebracht. GARTNER versteht darunter, analog zu DAYAL ET AL., die Einbindung von Funktionen, wie sie aus sozialen Onlinenetzwerken bekannt sind, in derzeit übliche BI-Systeme.<sup>254</sup>

Vor diesem Hintergrund stellt CBI für die Kombination kooperativer Mechanismen und Business Intelligence (in Abgrenzung zu den allgemeiner gefassten DSS) den derzeit gängigen Begriff dar. Er lässt sich gegen Social Business Intelligence (SBI) abgrenzen, welche die Extraktion und Nutzbarmachung von Daten aus sozialen Medien durch bestehende BI-Systeme fokussiert.<sup>255</sup>

### 3.3.1 Literaturüberblick

Die bisherigen Ausführungen weisen darauf hin, dass mit CBI ein Begriff existiert, der seine Wurzeln zwar in Gruppenentscheidungsprozessen hat, allerdings klar auf Business Intelligence als konzeptionelle und technologische Plattform fokussiert. Im Folgenden werden zur klaren Definition und Abgrenzung die derzeitigen Verständnisse des Begriffs und der zugehörige Stand der Forschung anhand einer Literaturrecherche erarbeitet. Darauf aufbauend können unterrepräsentierte Forschungsfelder identifiziert werden. Dieses als Review oder State-of-the-Art-Betrachtung bezeichnete Vorgehen ermöglicht es dann, die entsprechenden Felder unter Berücksichtigung existierender Ansätze zu füllen.<sup>256</sup>

Die Verwendung von Reviews ist in der wissenschaftlichen Literatur ausführlich dokumentiert, wobei die Autoren durchgängig darauf hinweisen, dass ein Review nicht nur die vorhandene Literatur ordnen sollte, sondern auch selbst einem ge-

<sup>252</sup> Vgl. Bitterer (2011); Bitterer (2012); Schlegel (2013); Schlegel (2014).

<sup>253</sup> Die folgenden Stadien sind „Peak of Inflated Expectations“, in dem die Aufmerksamkeit maximal ist und als überhöht gilt, „Through of Disillusionment“, in dem die Aufmerksamkeit auf Grund einer Rationalisierung stark nachlässt, „Slope of Enlightenment“, in dem die Technologie zur Reife kommt und wieder beachtet wird und „Plateau of Productivity“, in dem sich ein stabiler Zustand eingeschungen hat (vgl. Gartner Inc. (2013)).

<sup>254</sup> Vgl. Bitterer (2011); Abschnitt 3.2.3.

<sup>255</sup> Hauptzweck von SBI ist die Analyse von Daten aus sozialen Netzwerken (vgl. Dinter/Lorenz (2012): 3). Es beschreibt damit eine rein passive Nutzung sozialer Medien (vgl. Roe (2011); Rosemann et al. (2012)). Ein Verständnis von SBI im Sinne von DSS liegt aber z. B. bei SEITZ vor, der eine anonyme Plattform zur Unternehmenskennzahlenprognose vorschlägt (vgl. Seitz (2012): 28–30). Das Vorliegen einer ‚Gruppe‘ und damit eines GDSS-ähnlichen Ansatzes ist durch die Anonymität allerdings fraglich (vgl. Abschnitt 3.2.1; Abschnitt 3.2.2).

<sup>256</sup> Vgl. Mertens (1995): 45.

ordneten Prozess folgen muss.<sup>257</sup> Für die nachfolgend durchgeführte Recherche wird der Prozess nach FETTKE zu Grund gelegt (vgl. Abb. 22).<sup>258</sup>

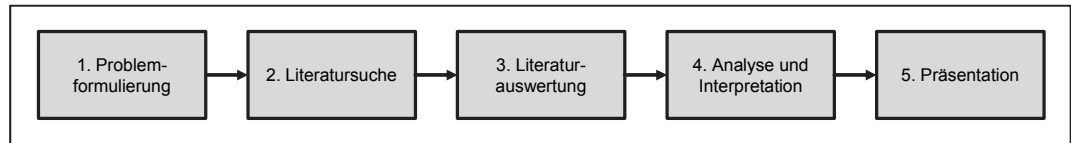


Abb. 22: Phasen der Reviewforschung

(Quelle: Fettke (2006): 260)

Die *Problemformulierung* wurde bereits zu Beginn dieses Abschnitts vorgenommen. Im Folgenden werden die Methodik bei der Literatursuche sowie die Auswertung der Ergebnisse vorgestellt. Daran anschließend wird ein Ordnungsrahmen definiert, der zur Analyse der Ergebnisse dient. Die daraus abgeleiteten unterschiedlichen Klassifikationsmöglichkeiten dienen nachfolgend zur Interpretation. Der vorliegende Abschnitt stellt dabei in Summe die Präsentation der Ergebnisse dar.

### 3.3.1.1 Recherchemethodik

Zur Einordnung der Ausrichtung der Literaturrecherche wird die Taxonomie von COOPER verwendet.<sup>259</sup> Innerhalb dieser werden sechs Charakteristika eine oder mehrere Ausprägungen zugeordnet (vgl. Abb. 23). Die für diese Arbeit zugeordneten Ausprägungen sind in der Abbildung grau hinterlegt.

<sup>257</sup> Vgl. Baker (2000): 220; Fettke (2006): 258; Brocke et al. (2009): 2207; Webster/Watson (2002): xv.

<sup>258</sup> Der Prozess nach FETTKE wird in der deutschsprachigen Wirtschaftsinformatik durch seine Veröffentlichung in der WIRTSCHAFTSINFORMATIK häufig als Basis genommen (vgl. Schief/Buxmann/Schiereck (2013): 422, Stroh/Winter/Wortmann (2011): 34).

<sup>259</sup> Vgl. Cooper (1988) zit. nach Cooper/Hedges (1994): 5.



Characteristics	Categories			
Focus	Research Outcomes	Research methods	Theories	Applications
Goal	Integration	Criticism	Central issues	
Perspective	Neutral representation		Espousal of position	
Coverage	Exhaustive	Exhaustive & Selective	Representative	Central or pivotal
Organization	Historical	Conceptual	Methodological	
Audience	Spezialized scholars	General scholars	Practitioners	General public

Abb. 23: Einordnung der Literaturrecherche in die Taxonomie von COOPER

(Quelle: Eigene Darstellung in Anlehnung an Cooper (1988))

Da der *Fokus* der Recherche auf dem Verständnis des Begriffs und der entsprechenden Definitionen liegt, beschäftigt sich das Review mit dem Aspekt der Theorie. Das entsprechende *Ziel* ist die integrative Darstellung vorhandener Arbeiten in einem Ordnungsrahmen sowie die Herausstellung zentraler Charakteristika und Problemstellungen. Dabei wurde als *Perspektive* eine neutrale Haltung gewählt. Die Recherche wurde mit dem Ziel einer umfangreichen *Abdeckung* des Publikationsspektrums durchgeführt. Sie fokussiert aber themenrelevante Zeitschriften und Konferenzen und ergänzt diese lediglich um weitere Quellen. Die *Organisation* der Ergebnisse und deren Präsentation finden auf Basis konzeptioneller Nähe statt; eine chronologische Reihenfolge wird aber, wo es sinnvoll ist, innerhalb dieser Struktur eingesetzt. Die Ergebnisse sind auf Grund ihrer breiten Darstellung ohne starke Verwendung von Fachtermini und auf Basis des beschriebenen Interesses von Marktforschungsunternehmen und Anwendern für ein *Publikum* aus (Fach-)Wissenschaftlern und Praktikern gleichermaßen geeignet.

Für die Literatursuche wurden zunächst Suchbegriffe festgelegt. Die Verwendung des bereits genannten Collaborative Decision Making ist dabei unvorteilhaft, da im Englischen – im Gegensatz zum Verständnis von GARTNER – damit nicht zwangsläufig ein systemischer Ansatz zur Entscheidungsunterstützungs/-findung

gemeint ist, sodass eine Suche mit Google Scholar<sup>260</sup> mehrere Tausend Treffer liefert, z. B. auch aus den Fachgebieten Medizin und Bildung.<sup>261</sup> Zur spezifischen Eingrenzung des Themenfeldes wurde daher auf ‚Collaborative Business Intelligence‘ als Suchbegriff abgestellt. Da ‚BI‘ die gängige Abkürzung des etablierten Feldes Business Intelligence ist, wurde zusätzlich auch der Begriff ‚Collaborative BI‘ als Suchbegriff verwendet.

Zur Durchführung der Recherche wurde ein zweistufiger Ansatz gewählt. Zunächst wurde eine Primärsuche in (wissenschaftlichen) Datenbanken und Zeitschriften durchgeführt. Die so gefundenen Artikel wurden in einem zweiten Schritt mit Hilfe einer Rückwärtssuche um weitere Publikationen ergänzt.<sup>262</sup>

Die Auswahl der zu durchsuchenden Zeitschriften und Datenbanken erfolgte auf Basis etablierter Rankings im Bereich der Wirtschaftsinformatik. Für den deutschsprachigen Raum ist die maßgeblich durch die Empfehlungen der *Wissenschaftlichen Kommission Wirtschaftsinformatik* (WKWI) in Abstimmung mit der *Fachgruppe Wirtschaftsinformatik der Gesellschaft für Informatik* (GI-FB WI) gegeben.<sup>263</sup> Die dort mit ‚A‘ klassifizierten, also höchst eingestuften Publikationsmedien wurden durch eine Suche in folgenden Datenbanken und Verlagsseiten abgedeckt:

- AIS Library<sup>264</sup>
- Google Scholar
- Palgrave<sup>265</sup>
- Sciencedirect<sup>266</sup>

---

<sup>260</sup> <http://scholar.google.com>. MEIER/CONKLING weisen darauf hin, dass Google Scholar – neben einer großen Menge an Publikationen diverser Forschungsbereiche – ca. 90% der seit 1990 erschienenen Publikationen aus dem ingenieurwissenschaftlichen Bereich beinhaltet (vgl. Meier/Conkling (2008)), was vor dem Hintergrund der gestaltungsorientierten WI von Interesse ist (vgl. Abschnitt 1.2.1).

<sup>261</sup> Beispiele sind Brunner (1998) und Sepucha et al. (2000). Außerhalb der wissenschaftlichen Literatur findet sich ‚Airport CDM‘, das von europäischen (<http://www.euro-cdm.org/>) und amerikanischen (<http://cdm.fly.faa.gov>) Flugsicherheitsbehörden als Bezeichnung für die Effizienzsteigerung durch die Zusammenarbeit aller am Prozess der Flugdurchführung und -vorbereitung Beteiligten genutzt wird.

<sup>262</sup> Eine Rückwärtssuche beschreibt die Identifikation relevanter Quellen auf Basis bereits bekannter Literatur. Eine Vorwärtssuche hingegen versucht Publikationen zu identifizieren, die ihrerseits auf bereits als relevant identifizierte Quellen verweisen und prüft diese auf Anwendbarkeit auf das Forschungsfeld (vgl. Webster/Watson (2002): xvi).

<sup>263</sup> Vgl. Wissenschaftliche Kommission Wirtschaftsinformatik (2008): 160–163.

<sup>264</sup> <https://ais.site-ym.com/?AISLibrary>.

<sup>265</sup> <http://www.palgrave-journals.com/pal/index.html>.

- Springerlink<sup>267</sup>
- TaylorFrancis<sup>268</sup>

Die so durchsuchten Medien beinhalten alle Zeitschriften und Beitragssammlungen des ‚Senior Scholars' Basket of Eight'<sup>269</sup>, der die wichtigsten Publikationen im anglo-amerikanischen Raum aufführt. In der Primärsuche wurden zudem – unabhängig von ihrer Zuordnung zu einem Publikationsmedium – alle weiteren Treffer der Datenbanken/Suchmaschinen berücksichtigt, um thematisch nahe Beiträge mit zu berücksichtigen. Tab. 4 stellt die Zuordnungen der durchsuchten Datenbanken und Verlagsseiten zu den Publikationen dar.

Datenbank/ Verlag	Abgedeckte Publikationen ( <i>kursiv</i> dargestellte Publikationen sind im 'basket of eight' enthalten)
<b>AIS</b>	<i>Information Systems Journal (ISJ)</i> <i>Journal of the Association of Information Systems (JAIS)</i> <i>Management Information Systems Quarterly (MISQ)</i> Proceedings of the European Conference on Information Systems (ECIS) Proceedings of the International Conference on Information Systems (ICIS)
<b>Google Scholar</b>	ACM Transactions Journals (ACMT) Communications of the Association for Computer Machinery (CACM) IEEE Software IEEE Transactions Journals (IEEE T) <i>Information Systems Research (ISR)</i> Informing Science Journal (INFSJ) <i>Journal of Management Information Systems (JMIS)</i> Management Science (MS) Organization Science (OS)
<b>Palgrave</b>	<i>European Journal of Information Systems (EJIS)</i> <i>Journal of Information Technology (JIT)</i>
<b>Sciencedirect</b>	I&O (Information and Organization) Information & Management (I&M) Information Systems (ISYS) Decision Support Systems (DSS) International Journal of Information Management (IJIM) <i>Journal of Strategic Information Systems (JSIS)</i>
<b>Springerlink</b>	Electronic Markets (EM) Wirtschaftsinformatik (WI)
<b>Taylor Francis</b>	Human-Computer Interaction (HCI)

Tab. 4: Zuordnung von Datenbanken und Verlagsseiten zu Publikationen  
(Quelle: Eigene Darstellung)

<sup>266</sup> <http://www.sciencedirect.com>.

<sup>267</sup> <http://www.springerlink.com>.

<sup>268</sup> <http://www.tandfonline.com>.

<sup>269</sup> Vgl. Association for Information Systems (2011).

Da das Ziel des Reviews ein Überblick über das allgemeine aktuelle Verständnis und den Stand der Forschung ist, wurde unter Berücksichtigung der Annahme von WITTGENSTEIN ET AL., dass die Bedeutung eines Wortes sich über die Verwendung in der Sprache definiert,<sup>270</sup> zusätzlich eine Web-Suche mit Google<sup>271</sup> durchgeführt. Dabei wurden jeweils die ersten 100 Treffer berücksichtigt und aus diesen die nicht redaktionell geführten Webseiten ausgeschlossen. Anbieterpräsentationen wurden nicht berücksichtigt, um einen neutralen Standpunkt zu gewährleisten und keine verkaufsorientierten Beschreibungen in die Recherche aufzunehmen. Als Suchbegriff wurde für alle Suchen die feste Wortfolge verwendet. Durchsucht wurden Titel, Zusammenfassungen (Abstracts) und Volltexte, sofern diese Möglichkeit bestand.

### 3.3.1.2 Rechercheergebnisse

Die Primärsuche<sup>272</sup> führte zu 402 Treffern in wissenschaftlichen Datenbanken, von denen 36 für ein detailliertes Review ausgewählt wurden, was einer Quote von ca. 9,0% entspricht. Die Quote für den Begriff ‚Collaborative Business Intelligence‘ lag dabei mit 14,9% erheblich höher als die von ‚Collaborative BI‘ mit 4,1%. Die Websuche führte zu neun weiteren Beiträgen (von den jeweils ersten 100 Treffern der Suche). Eine detaillierte Aufschlüsselung ist in Tab. 5 dargestellt.

	Collaborative Business Intelligence		Collaborative BI	
	Treffer	verwendet	Treffer	verwendet
<b>Datenbanken/ Verlage</b>	181	27 ≈ 14,9%	221	9 ≈ 4,1%
<b>Google Websuche</b>	100 (von 80.200)	6 = 6,0%	100 (von 16.900)	3 = 3,0%

Tab. 5: Anzahl Treffer und verwendeter Beiträge der Literatursuche

(Quelle: Eigene Darstellung)

Für die geringe Quote, insbesondere in Bezug auf den Suchbegriff ‚Collaborative BI‘, lassen sich mehrere Gründe identifizieren:

1. Die Verwendung unterschiedlicher Datenbanken führte zur Mehrfachfindung einzelner Artikel. So listet Google Scholar z. B. auch einzelne (aber nicht alle) Artikel, die durch die AIS-Suchmaschine abgedeckt werden.

<sup>270</sup> Vgl. Wittgenstein et al. (2009): 25.

<sup>271</sup> <http://www.google.com>.

<sup>272</sup> Stichtag der Suche war der 27.01.2015.

2. Beiträge, die weder in deutscher, noch in englischer Sprache vorlagen, wurden nicht berücksichtigt.
3. Unterschiedliche Treffer, die einzelne Kapitel einer Monographie darstellten, wurden zusammengefasst und nur als ein verwendeter Artikel gezählt.
4. Die Suche nach ‚Collaborative BI‘ führte zu Treffern, bei denen BI nicht als Abkürzung verwendet wurde, sondern als Bestandteil eines nachfolgenden Wortes.<sup>273</sup> Derart gefundene, themenfremde Artikel wurden nicht berücksichtigt. Auch wurde ‚collaborative‘ nicht immer in Verbindung mit BI, bzw. Business Intelligence, als feststehender Term verwendet, sondern auch als unspezifisches Attribut, sodass keine Definition im eigentlichen Sinne gegeben war.

Die durch die Primärsuche identifizierten Artikel wurden durch die Rückwärtssuche auf 51 Publikationen erhöht. Die Darstellung der verwendeten Publikationen nach Jahren zeigt dabei einen erkennbaren Sprung in der Verwendung des Begriffes mit Beginn des Jahres 2010 (vgl. Abb. 24). Die Mehrzahl aller Publikationen ist in den Jahren 2010–2014 vorzufinden, was auf ein steigendes Interesse, zumindest aber eine steigende Verwendung des Begriffes, hinweist.

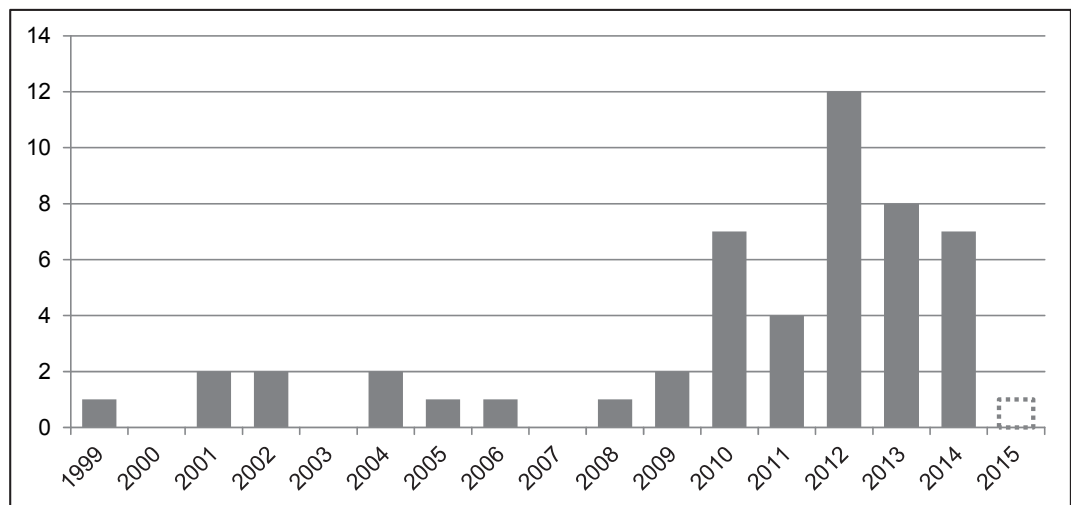


Abb. 24: Anzahl Publikationen zu CBI nach Jahren

(Quelle: Eigene Darstellung)

<sup>273</sup> Als Beispiel sei MORIN aufgeführt: “The project, a collaborative, bi-national effort of more than 30 major botanical institutions in the United States and Canada, draws on the expertise of botanists around the world to write and review treatments.” (Morin (1997): 338).

### 3.3.2 Aufbau eines Ordnungsrahmens

Zur Klassifikation der gefundenen Publikationen wird ein zweidimensionaler Ordnungsrahmen verwendet, der sich aus dem üblichen BI-Prozess ableitet und zwischen Datenbeschaffung/-extraktion/-bereitstellung (im Sinne der Bereitstellung für das zentrale DWH) und Datenauswertung/-analyse unterscheidet.<sup>274</sup> Beide Aspekte lassen sich, ausgehend von dem ‚klassischen‘ Ansatz, durch Kooperation erweitern, indem andere als die üblichen Quellen und Analysten mit eingebunden werden.

SCHLAGWEIN/SCHODER/FISCHBACH unterscheiden für Informationsressourcen verschiedene Formen der Öffnung, die von einer internen/exklusiven Verwendung über einen gruppenorientierten Ansatz hin zu einer offenen, externen Nutzung reichen.<sup>275</sup> Angewendet auf BI-Systeme ergibt sich die Möglichkeit, sowohl den Quellenumfang (Datenbereitstellung) als auch den Analystenumfang (Datenauswertung) zu erhöhen.

Die nicht-kooperative Variante ist damit die Nutzung bestehender Systeme und Auswertung nur im Kreise der (spezialisierten) Abteilung. Eine Erweiterung findet statt, sobald zunächst das eigene Unternehmen, d. h. weitere Abteilungen mit einbezogen werden. Dies kann z. B. im Rahmen von unternehmensinternen Prognosen geschehen oder durch den Einsatz kontextsensitiver Informationen (Abteilungs-/Personenspezifika), wie es durch den verstärkten Einsatz mobiler Entscheidungsunterstützungslösungen („Mobile BI“) zunehmend zum Tragen kommt.<sup>276</sup> Für den Miteinbezug anderer Abteilungen in den Analyseprozess können Konzepte wie Annotationen und Verlinkung von Entscheidungen und verwendeten Daten eingesetzt werden.<sup>277</sup>

Sofern zwischen unabhängigen Unternehmen ein entsprechendes Vertrauensverhältnis etabliert ist, kommt eine kooperative Datenhaltung und/oder Analyse in Betracht.<sup>278</sup> Im Gegensatz zur reinen Erweiterung der unternehmenseigenen Datenbasis (keine Kooperation) findet dann ein netzwerkbasierter Austausch von

---

<sup>274</sup> Vgl. Abschnitt 3.1.2.

<sup>275</sup> Vgl. Schlagwein/Schoder/Fischbach (2010): 5.

<sup>276</sup> Vgl. Seitz (2012): 28–30; Chen/Chiang/Storey (2012): 1168, 1178.

<sup>277</sup> Vgl. Turban/Liang/Wu (2011): 146.

<sup>278</sup> Vgl. Liu/Daniels (2012): 4.

Daten statt.<sup>279</sup> Eine gemeinschaftliche Analyse ist z. B. in der Versicherungswirtschaft bei Risikoanalysen denkbar.<sup>280</sup>

Vollständig externe Datenquellen, die nicht als reine Hinzuerfassung in bestehenden Systeme abgebildet werden, können im Rahmen von öffentlich zugänglichen Informationsdatenbanken vorgefunden werden. Seit 2009 wird für diese verstärkt der Begriff *open data* verwendet.<sup>281</sup> Eine beispielhafte Öffnung hin zu externen Analysten findet sich im Bereich *Crowdsourcing*, d. h. der Aufgabenverlagerung an eine (anonyme) Menge von Personen.<sup>282</sup>

Tab. 6 stellt die für den Ordnungsrahmen definierten Ausprägungen von Kooperation und die entsprechenden Erläuterungen zusammenfassend dar.

Der durch die beschriebenen Dimensionen aufgespannte Ordnungsrahmen ist im Anschluss in Abb. 25 dargestellt. Eingetragen sind die Publikationen, die durch die Literaturrecherche identifiziert wurden. Dabei wurden Publikationen mit identischem oder nahezu identischem Begriffsverständnis durch rechteckige Umfassungen zusammengefasst. Tab. 7 stellt die Zuordnung der im Ordnungsrahmen eingetragenen Nummern zu den eigentlichen Artikeln dar.

---

<sup>279</sup> Beispiele hierfür finden sich bei strategischen Allianzen (vgl. Rizzi (2012): 187) und Regierungsorganisationen (vgl. Martins/daCosta/deSousa (2012): 99).

<sup>280</sup> Ein Beispiel hierfür liefern LIU/DANIELS mit dem System FISH (Fraud Information System Holland) (vgl. Liu/Daniels (2012): 4). Die zuständige Stiftung Zentrales Informationssystem (CIS) speichert Informationen über Versicherungs- und Betrugsfälle (vgl. Stichting CIS (2015)). Damit in Deutschland vergleichbar ist die SCHUFA (<https://www.schufa.de/>), die eine zentrale Bonitätsprüfungsstelle angeschlossener Institutionen und Unternehmen der Finanz- und Versicherungsbranche darstellt.

<sup>281</sup> Vgl. Alanazi/Chatfield (2013): 1. Die Bereitstellung von ‚open data‘ seitens Regierungsorganisationen wird im angelsächsischen Raum stärker forciert, vgl. z. B. <http://www.data.dov> (US) oder <http://www.data.gov.uk> (GB). Das deutsche Pendant ist seit Februar 2013 unter <http://www.govdata.de> erreichbar. Neben öffentlichen Stellen, Behörden und Statistischen Ämtern kommen Nicht-Regierungsorganisationen und Unternehmen als Datenbereitsteller in Frage.

<sup>282</sup> Vgl. Böhringer et al. (2010): 5. Der von den Autoren verwendete Begriff ‚BI crowdsourcing‘ hat derzeit keine wissenschaftliche Relevanz. Eine Suche mit Google Scholar führt zu nur einem Treffer (eine Masterarbeit an der Universidade de Brasilia), die eine (portugiesischsprachige) Referenz auf BÖHRINGER ET AL. vornimmt, ohne das Konzept selbst aufzugreifen (vgl. Martins (2013): 10).

Dimension	Ausprägungen	Erläuterung
<b>Beteiligte (Quellen) bei der Datenbereitstellung</b>	Bestehende Systeme	<ul style="list-style-type: none"> <li>• Verwendung ‚klassischer‘ Quellen (z. B. ERP, PPS)</li> <li>• Zusätzliche Nutzung bewusst hinzuerfasster, ggf. redaktionell aufbereiteter Quellen (z. B. Marktanalysen)</li> </ul>
	Eigenes Unternehmen	<ul style="list-style-type: none"> <li>• Verwendung zusätzlicher, unternehmensinterner Quellen</li> <li>• Einbezug von Mitarbeiterwissen (z. B. zu Prognosezwecken)</li> </ul>
	Partner / Netzwerke / Assoziierte	<ul style="list-style-type: none"> <li>• Einbezug von Partnerunternehmen / Netzwerkmitgliedern aus strategischen Allianzen (z. B. Einkaufs-genossenschaften, Allianzen in der Luftfahrt) oder verbundenen Gesellschaften (z. B. Sparkassen, Genossenschaftsbanken)</li> <li>• ‚Geben und nehmen‘-Prinzip</li> </ul>
	Externe / Unabhängige	<ul style="list-style-type: none"> <li>• Einbezug unabhängiger Datenbereitsteller (z. B. Behörden, ‚open data‘-Projekte)</li> <li>• Verwendung indirekter Lieferanten (z. B. durch Auswertung sozialer Netzwerke und Nachrichtendienste wie twitter) insbesondere zur Meinungserforschung</li> </ul>
<b>Beteiligte (Analysten) bei der Datenauswertung</b>	Fach- / Analyseabteilungen	<ul style="list-style-type: none"> <li>• Hauptverantwortung liegt in Fachabteilung</li> <li>• Analyseaufgaben können zentralisiert ausgeführt werden (Competency Center)</li> </ul>
	Eigenes Unternehmen	<ul style="list-style-type: none"> <li>• Rückgriff der Fachabteilungen auf interne Experten</li> <li>• (Technische) Unterstützung des Prozesses (z. B. durch Annotations-/Zuordnungsfunktionen des Berichtswesens)</li> </ul>
	Partner / Netzwerke / Assoziierte	<ul style="list-style-type: none"> <li>• Einbezug von Analysten verbundener Unternehmen (z. B. Branchenexperten, Spezialisten einer Genossenschaft)</li> </ul>
	Externe / Unabhängige	<ul style="list-style-type: none"> <li>• Einbezug unabhängiger Dritter</li> <li>• Unter Berücksichtigung von Social Media die Verlagerung in die ‚Crowd‘, d. h. eine Menge anonymer Individuen (z. B. für Markteinschätzungen oder Prognosen)</li> </ul>

Tab. 6: Dimensionen und Ausprägungen des Ordnungsrahmens für Collaborative Business Intelligence

(Quelle: Eigene Darstellung)



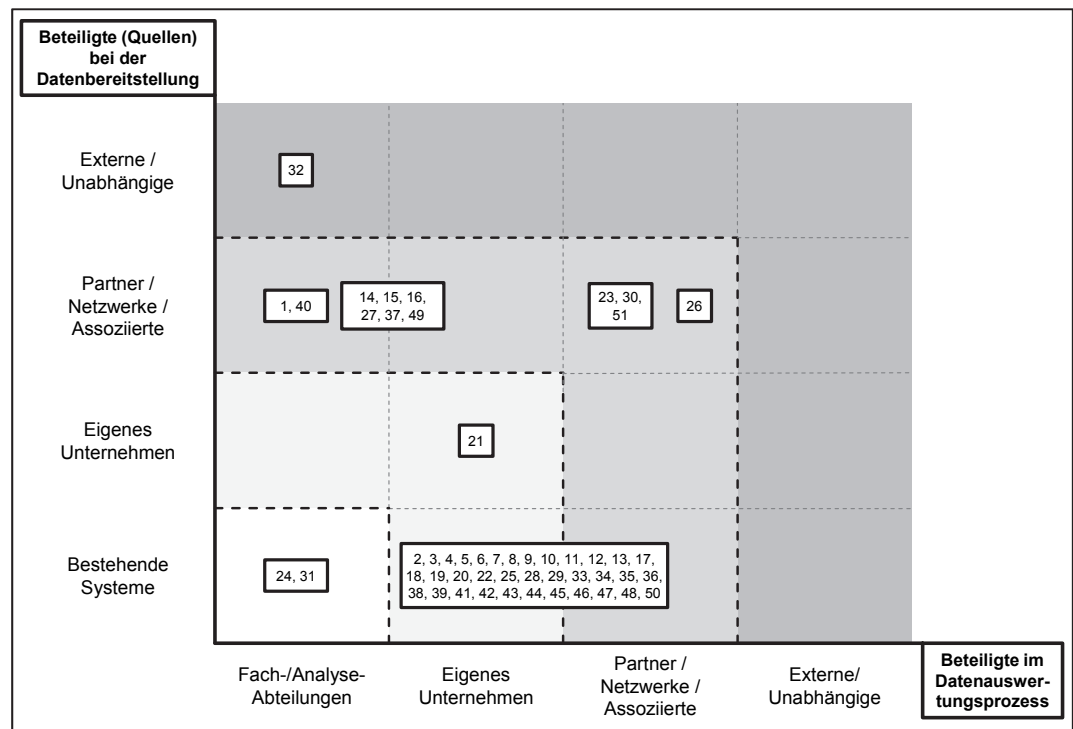


Abb. 25: Ordnungsrahmen für Collaborative Business Intelligence

(Quelle: Eigene Darstellung)

# Quelle	# Quelle	# Quelle
1 Alwis (2004)	18 Hägele (2015)	35 Panja/Paul (2014)
2 Audimoolam/Dutta (2005)	19 Hagerty/Sallam/Richardson (2012)	36 Paul (2014)
3 Averbuch (2013)	20 Hammergren/Simon (2009)	37 Rizzi (2012)
4 Berthold et al. (2010)	21 Hammond (1999)	38 Rouse (2012)
5 Bitterer (2011)	22 Horakova/Skalska (2013)	39 Schlegel (2014)
6 Collins (2001)	23 Klarmann/Roth/Franczyk (2013)	40 Schwalm/Bange (2004)
7 Dayal et al. (2008)	24 Ladley (2010)	41 Simon/Shaffer (2002)
8 Dayal et al. (2010)	25 Lipinski (2015)	42 Stavrianos/Henderson (2006)
9 Devlin (2012)	26 Liu/Daniels (2012)	43 The Data Warehousing Institute (2013)
10 Dresner (2013)	27 Martins/daCosta/deSousa (2012)	44 Teruel et al. (2014)
11 Dyche (2012)	28 McDonald (2002)	45 The Law Dictionary (2013)
12 EMA Inc. (2012)	29 McKnight (2014)	46 Thierauf (2001)
13 Golfarelli (2009)	30 Mettler/Raber (2011)	47 Trujillo/Maté (2012)
14 Golfarelli et al. (2010)	31 Molensky et al. (2010)	48 Turban/Liang/Wu (2011)
15 Golfarelli et al. (2011)	32 Muntean (2012b)	49 Vera-Baquero/Colomo-Palacios/Molloy (2013)
16 Golfarelli et al. (2012a)	33 Muntean (2012a)	50 Weber/Kolb (2014)
17 Gurjar/Rathore (2013)	34 Nelson (2010)	51 Werner et al. (2010)

Tab. 7: Zuordnung der Publikationen zum Ordnungsrahmen für Collaborative Business Intelligence

(Quelle: Eigene Darstellung)

Erkennbar ist, dass ein Großteil der Publikationen ein ähnliches Verständnis im Bereich ‚Nutzung bestehender Systeme zur Datengenerierung‘/‘Beteiligung unternehmensinterner (und teilweise partnerschaftlicher externer) Analysten‘ aufweist. Ebenfalls verbreitet ist das Verständnis von CBI im Bereich gemeinschaftlicher Datenhaltung und – wenn auch geringer – im Bereich partnerschaftlicher Analyse. Im Rahmen dieser Arbeit gilt daher folgende weit gefasste Definition:

**Definition (5)** *Collaborative Business Intelligence* beschreibt die bewusste Erweiterung des Analystenkreises im Business-Intelligence-Prozess um unternehmensinterne und -externe Partner sowie die dafür notwendige Öffnung der Datenbestände der Beteiligten.

Lediglich drei Beiträge weisen ein vollständig abweichendes Begriffsverständnis auf und beschreiben CBI als Analyse von Kooperationsmechanismen in Unternehmen<sup>283</sup>, als den Einsatz spezialisierter Software-Agenten<sup>284</sup> oder als Social BI, gepaart mit Location Intelligence.<sup>285</sup> Bemerkenswert ist, dass die Definition von MUNTEAN erheblich von der in einer anderen Arbeit der Autorin aus dem gleichen Jahr abweicht.<sup>286</sup> Die Dynamik in der Begriffsfindung wird auch von GOLFARELLI verdeutlicht, dessen ursprüngliche Definition<sup>287</sup> in neueren Arbeiten nicht mehr vorzufinden ist, sondern den Aspekt unternehmensübergreifender BI betont.<sup>288</sup>

Initial können drei Hauptgebiete des Verständnisses von CBI identifiziert werden.<sup>289</sup>

- Interne Kommunikation / Internal Communication (IC)
- Partnerschaftliche Datenhaltung / Partnership in data (PD)
- Partnerschaftliche Analyse / Partnership in analysis (PA)

<sup>283</sup> Vgl. Ladley (2010): 270.

<sup>284</sup> Vgl. Molensky et al. (2010): 127.

<sup>285</sup> Vgl. Muntean (2012b): 4. MUNTEAN versteht unter ‚Location Intelligence‘ die Verknüpfung von Geschäftsdaten mit geografischen Informationen.

<sup>286</sup> Vgl. Muntean (2012a): 196.

<sup>287</sup> Vgl. Golfarelli (2009): 293.

<sup>288</sup> Vgl. Golfarelli et al. (2010): 37, 39; Golfarelli et al. (2011): 244f.; Golfarelli et al. (2012b): 393.

<sup>289</sup> Die Aufteilung folgt einer früher durchgeführten und publizierten Literaturübersicht und führt daher neben den deutschen Bezeichnungen auch die englischen Bezeichnungen, die als Grundlage für die Abkürzungen der einzelnen Felder dienen (vgl. Kaufmann/Chamoni (2014): 3742f.).

Diese Aufteilung greift auch auf die zentralen Kriterien der unternehmensübergreifenden Kooperation zurück, insbesondere auf die Intensität und die Anzahl der Partner. Während die interne Kommunikation lediglich einen Austausch von Informationen und Ereignissen bedeutet, wird bei einer gemeinsamen Datenhaltung bereits ein arbeitsteiliges Vorgehen verwendet, was bei einer partnerschaftlichen Datenanalyse zu einem gemeinschaftlichen Vorgehen führen kann.<sup>290</sup> Der folgende Abschnitt stellt die wesentlichen Aspekte dieser Gebiete sowie die Ausrichtung der durch die Recherche gefundenen Beiträge vor.

### 3.3.3 Eigenschaften unterschiedlicher Formen

Zur Darstellung der inhaltlichen Fokussierung der Beiträge wird ein Raster verwendet, das der in der Business Intelligence etablierten Aufteilung in fachliche, organisatorische und technologische Aspekte<sup>291</sup> folgt. TURBAN/LIANG/WU schlagen für die Einteilung potenzieller Forschungsfelder im Bereich Collaborative Decision Making eine ähnliche Struktur vor, ergänzen jedoch noch den Aspekt der ökonomischen Fragestellungen, insbesondere auch im Sinne der Erfolgsmessung und der Bildung von Anwendungsfällen.<sup>292</sup> Auf die vier in der Art identifizierten Kriterien wird im weiteren Verlauf als *TOFE-Klassifikation* referenziert, abgeleitet aus der Beschreibung für technische (T), organisatorische (O), fachliche (F) und ökonomische („economic“, E) Aspekte. Tab. 8 zeigt eine Einschätzung zur Verteilung der Beiträge in den jeweiligen Schwerpunkten nach den genannten Aspekten. Dargestellt ist, mit welchem Fokus das Thema behandelt wurde oder, sofern der Beitrag nur die Begriffsfindung unterstützt, welcher Aspekt dafür als maßgeblich identifiziert wurde. Beiträge, die mehrere Aspekte behandeln, werden auch mehrfach aufgeführt.

Schwerpunkt	T	O	F	E
<b>Interne Kommunikation/ Internal Communication (IC)</b>	33	7	2	6
<b>Partnerschaftliche Datenhaltung/ Partnership in data (PD)</b>	7	1	5	6
<b>Partnerschaftliche Analyse/ Partnership in analysis (PA)</b>	3	1	0	4

Tab. 8: Anzahl der Beiträge in der Zuordnung zur TOFE-Klassifikation

(Quelle: Eigene Darstellung)

<sup>290</sup> Vgl. die Ausprägungen im morphologischen Kasten (Abb. 7).

<sup>291</sup> Vgl. Böhringer et al. (2010): 8; Chamoni/Gluchowski (2004): 120; Gansor/Totok/Stock (2010): 39–42.

<sup>292</sup> Vgl. Turban/Liang/Wu (2011): 151.

### 3.3.3.1 Interne Kommunikation

Die Mehrheit der untersuchten Beiträge folgt bei der Definition und Beschreibung des Begriffes einem technischen Verständnis. 33 der insgesamt 36 als CBI-IC klassifizierten Publikationen beschreiben entweder CBI als technische Entwicklung oder leiten ihr Verständnis von technischen Entwicklungen und Gegebenheiten ab. Beiträge, die vor dem Jahr 2008 veröffentlicht wurden, sind dabei in einem Umfeld entstanden, das nicht in dem Maße von der Verbreitung sozialer Netzwerke geprägt war, wie es seitdem der Fall ist.<sup>293</sup> Die ersten Publikationen fokussieren daher bestehende Portalkonzepte, Workflow-Komponenten sowie Wissensmanagement-Systeme und beschreiben eine Erweiterung der entsprechenden Konzepte um Echtzeitkommunikation und gemeinsame Datenhaltung.<sup>294</sup> Die Ansätze sind dabei häufig produktorientiert und -getrieben und beziehen sich vornehmlich auf die Präsentationsebene von BI-Systemen.<sup>295</sup>

Auch die weiteren von technischem Verständnis geprägten Arbeiten fokussieren diese Ebene und betonen die Möglichkeiten der Erweiterung von BI-Systemen um Funktionalitäten, wie sie aus sozialen Medien bekannt sind. Erwähnt werden die Annotation von Berichtselementen, direkte Feedback-Möglichkeiten durch Verantwortungszuordnungen und Nachrichtendienste sowie einfache Benutzeroberflächen.<sup>296</sup> Eine exemplarische Definition findet sich bei MUNTEAN: „Collaborative Business Intelligence – the integration of information sharing features and functionality of popular Web 2.0 technologies and social media platforms within a BI platform (...)“<sup>297</sup>.

DEVLIN bezieht sich in seiner Ausarbeitung auf den organisatorischen Aspekt und beschreibt CBI als das Einbeziehen mehrerer Teammitglieder in einen Entschei-

<sup>293</sup> RICHTER/RIEMER/BROCKE weisen in ihrer Literaturrecherche darauf hin, dass ein Großteil der Forschung zu OSNs erst in den Jahren 2008 und 2009 publiziert wurde (vgl. Richter/Riemer/Brocke (2011): 92). Die durchschnittliche Nutzung von sozialen Netzwerken durch Privatpersonen lag 2005 bei ca. 8%, 2008 bei ca. 30% und nach 2010 bei über 60% (vgl. Pew-Research (2013)).

<sup>294</sup> Vgl. Collins (2001): 185; Simon/Shaffer (2002): 162; Stavrianos/Henderson (2006); Thierauf (2001): 103.

<sup>295</sup> Vgl. Hammond (1999): 117; McDonald (2002): 30.

<sup>296</sup> Vgl. Audimoolam/Dutta (2005): 9; Averbuch (2013); Bitterer (2011): 11; Dyche (2012); Golfarelli (2009): 293; Gurjar/Rathore (2013): 83; Hagerty/Sallam/Richardson (2012): 22; Hägele (2015): 333; Hammergren/Simon (2009): 311-316; Horakova/Skalska (2013): 54; Lipinski (2015); McKnight (2014): 165; Nelson (2010): 8; Panja/Paul (2014): 112f.; Paul (2014): 69; Rouse (2012); Schlegel (2014); The Data Warehousing Institute (2013); Teruel et al. (2014): 424-427; Trujillo/Maté (2012): 113; Weber/Kolb (2014): 20f.

<sup>297</sup> Muntean (2012a): 196.

dungsprozess. Er nimmt dabei keinen direkten Bezug auf die technische Unterstützung. Sein Verständnis ist davon geprägt, dass die Analyse von Daten weiterhin zunächst von einer Person durchgeführt wird und erst im Weiteren geteilt und diskutiert wird – eine gleichzeitige, tatsächlich gemeinschaftliche Analyse liegt nicht vor.<sup>298</sup>

Unter den identifizierten Publikationen beschäftigen sich zwei mit einem ganzheitlichen Ansatz, wenngleich das Kernverständnis auch bei diesen von der technischen Möglichkeit und deren Umsetzung abgeleitet wird. DAYAL ET AL. verwenden eine Praxisstudie in einem Datenzentrum, um zu demonstrieren, welche Herausforderungen bei dem Einsatz einer kooperativen Umgebung im Bereich BI auftreten. Der konkrete Anwendungsfall besteht dabei aus der Beobachtung mehrerer Performance-Indikatoren des Datenzentrums, die mit möglichst geringer Latenzzeit von mehreren Personen ausgewertet werden sollen. Um dies technisch zu unterstützen, lösen sich die Autoren vom klassischen DWH-Schema und verzichten auf die Integration aller Daten in eine zentrale Datenbank. Sie verwenden stattdessen eine zusätzliche Ebene, die eine ganzheitliche Sichtweise auf die Daten ermöglicht, ohne diese selbst zu verändern („ontology and meta data layer“<sup>299</sup>).<sup>300</sup> In diesem Zusammenhang werden auch organisatorische Aspekte diskutiert.<sup>301</sup> Die Schwierigkeiten der automatischen Generierung von Ontologien<sup>302</sup> zur Auswertung werden adressiert, allerdings nicht im Rahmen der Arbeit gelöst. Die ökonomische Rechtfertigung basiert auf der Schaffung einer Echtzeitfähigen Analyseumgebung, die operative Systeme stärker an existierende BI-Lösungen bindet und so zu schnelleren – und damit wirtschaftlich besseren – Entscheidungen führen soll.<sup>303</sup>

---

<sup>298</sup> Vgl. Devlin (2012): 16.

<sup>299</sup> Dayal et al. (2008): 14.

<sup>300</sup> Vgl. Dayal et al. (2008): 8–19.

<sup>301</sup> Vgl. Dayal et al. (2008): 20–23.

<sup>302</sup> Der aus dem Griechischen stammende Begriff *Ontologie* (engl.: ontology) wurde und wird in der Philosophie seit mehreren Jahrhunderten für die ‚das Sein‘ behandelnde Wissenschaft verwendet. In der modernen Informatik steht die Ontologie konkreter für ein formales Repräsentationssystem von Begriffen und deren Beziehungen in einem fest umschriebenen Kontext – einer *Domäne*. Ontologien ermöglichen die Aufstellung formal abbildbarer, maschinell erfassbarer Beziehungssysteme – unter anderem zur Durchführung automatisierter Validierungen und Schlussfolgerungen (vgl. Busse et al. (2014): 286f.; Furrer (2014): 308–310). Die hierarchische Abbildung einer Dimension in einem DWH (vgl. Abschnitt 3.1.3) kommt damit der Idee einer Ontologie im Rahmen der BI nahe. Konzepte und Methoden zu Ontologien können daher auch im Rahmen von DWH-Problematiken angewendet werden (vgl. Abschnitt 4.3.3).

<sup>303</sup> Vgl. Dayal et al. (2008): 9, 12f.

Ebenfalls auf der Schaffung von zusätzlichen Architekturebenen basiert das Konzept von BERTHOLD ET AL. Die Autoren unterscheiden dabei zwischen „infospaces“<sup>304</sup>, die die konzeptionellen Informationen über die Daten aufnehmen und „dataspaces“<sup>305</sup>, die die eigentlichen Daten vorhalten. Organisatorische und fachliche Aspekte werden im Zusammenhang der wirtschaftlichen Vorteile diskutiert. BERTHOLD ET AL. stellen dabei auf das Prinzip der Informationsselbstversorgung<sup>306</sup> ab und argumentieren, dass die Befähigung von Fachanwendern zur eigenverantwortlichen Erstellung von Berichten zu einer Reduktion von Redundanz der Arbeit von Fachbereichen und IT-Abteilungen führt und die Gesamtkosten für die Unternehmung gesenkt werden können.<sup>307</sup>

Die einzige empirische Arbeit zur Wirtschaftlichkeitsbetrachtung von CBI-IC liefern LIU/KIM/SUN. Sie untersuchen den Zusammenhang zwischen dem Einsatz von kooperativen Mechanismen im BI-Umfeld und einer ‚guten‘ Entscheidung. Die Autoren finden diesen bestätigt, weisen jedoch auf die begrenzte Aussagefähigkeit der Studie hin, da die Teilnehmer ausschließlich chinesische Studierende waren.<sup>308</sup>

Als CBI-IC werden damit abschließend alle Beiträge klassifiziert, die die Quellen für den Analyseprozess auf bestehende und unternehmenseigene Systeme und Ressourcen einschränken. Die Literaturrecherche zeigt, dass entsprechende Beiträge auch davon ausgehen, dass der Kreis der Beteiligten auf das eigene Unternehmen oder explizite Partnerunternehmen eingegrenzt ist.

### **3.3.3.2 Gemeinschaftliche Datenhaltung**

Der Einbezug unternehmensfremder Quellen ist gegeben, wenn externe Daten mit verwendet werden, das betrachtete Unternehmen also als reiner Konsument auftritt, oder Daten für externe Analyse anderer bereitgestellt werden. Ohne eine Veränderung der klassischen Analysetätigkeit zu fordern, beschreiben

---

<sup>304</sup> Berthold et al. (2010): 4.

<sup>305</sup> Ebd.

<sup>306</sup> Das Konzept der Informationsselbstversorgung wird als ‚self-service BI‘ in der Literatur mehrfach genannt und beschreibt die Möglichkeit der Endanwender von BI-Systemen, nicht nur vorgefertigte Analysen oder Berichte zu konsumieren, sondern selbst konzeptionell tätig zu werden und Datenmodelle, Berichte, Analysen und weitere Elemente der Systeme zu erstellen und zu ändern (vgl. Baars et al. (2014): 16; Geist/Kluin/Ritz (2013): 48–51; Imhoff/White (2011)).

<sup>307</sup> Vgl. Berthold et al. (2010): 5.

<sup>308</sup> Vgl. Liu/Kim/Sun (2012): 7f.

SCHWALM/BANGE eine technisch motivierte Lösung zum Einbezug von strukturierten, unternehmensexternen Daten mit Hilfe einer Standardisierung auf Basis der *Extensible Markup Language* (XML).<sup>309</sup> Als rein betriebswirtschaftlich motivierte Definition kann das Verständnis von ALWIS gelten, der CBI als „external intelligence“<sup>310</sup> von Erkenntnissen der eigenen Unternehmung abgrenzt. Dieses externe Wissen soll durch die Kooperation mit anderen Unternehmen und die Einbindung fremder Informationen geschaffen werden.<sup>311</sup> Beide beschriebenen Ansätze beschränken sich auf eine Anreicherung des bestehenden Analyseprozesses der Fachabteilungen.

Ein Miteinbezug weiterer Unternehmensteile ist in den Beiträgen von GOLFARELLI ET AL. und RIZZI erkennbar.<sup>312</sup> Die Autoren beschreiben das Konzept eines P2P-basierten *Business Intelligence Network* (BIN), das einer gemeinsamen Datenhaltung multidimensionaler Daten mehrerer Unternehmen oder Unternehmensteile dient.<sup>313</sup> Neben der Beschreibung einer technischen Lösung werden die Thematik der fachlichen Übereinstimmung von Daten verschiedener DWH-Systeme sowie organisatorische Fragestellungen zum Umgang der Beteiligten miteinander diskutiert. Dabei wird auch das Fehlen von geeigneten Messgrößen für die Übereinstimmung und Verwendbarkeit der Datenschemata der partizipierenden Unternehmen adressiert und als notwendiger Forschungsbedarf identifiziert.<sup>314</sup> Die ökonomische Rechtfertigung sehen die Autoren in der ansonsten nicht möglichen Verwendung unternehmensexterner Informationen im Kontext der Entscheidungsfindung.<sup>315</sup>

Ein technisch getriebener Ansatz, der auch fachliche Fragestellungen behandelt, von MARTINS/DACOSTA/DESOUSA vorgestellt. Mit Referenz auf die Arbeit von BERTHOLD ET AL. entwickeln die Autoren eine Referenzarchitektur, die auf einem verteilten Ansatz fußt, der ebenfalls eine zentrale Ontologie vorsieht, die die Da-

---

<sup>309</sup> Vgl. Schwalm/Bange (2004): 12f.

<sup>310</sup> Alwis (2004): 45.

<sup>311</sup> Vgl. Alwis (2004): 45.

<sup>312</sup> Der Beitrag von VERA-BAQUERO/COLOMO-PALACIOS/MOLLOY (vgl. die Einordnung in Abb. 25) liefert keine eigenständige Definition, sondern folgt dem Verständnis von RIZZI (vgl. Vera-Baquero/Colomo-Palacios/Molloy (2013): 31).

<sup>313</sup> Vgl. Golfarelli et al. (2012a): 147f.; Golfarelli et al. (2011): 244f.; Golfarelli et al. (2010): 37f.; Rizzi (2012): 194. Das Konzept P2P-basierter Netzwerke im Bereich der BI wird im Fortgang dieser Arbeit detaillierter vorgestellt (vgl. Kapitel 4).

<sup>314</sup> Vgl. Rizzi (2012): 202.

<sup>315</sup> Vgl. Golfarelli et al. (2012a): 147; Golfarelli et al. (2012b): 393.



ten mehrerer beteiligter DWH-Systeme fachlich vereint.<sup>316</sup> Die Entwicklung der Ontologie ist dabei nicht Gegenstand des Beitrags. Als Anwendungsbeispiel wird die Integration von BI-Systemen mehrerer Ministerien der brasilianischen Regierung genannt.<sup>317</sup>

### 3.3.3.3 *Gemeinschaftliche Analyse*

Vier der 51 betrachteten Beiträge erweitern sowohl die Datenhaltung, bzw. den Datenbeschaffungsprozess, als auch den Analyseprozess über die Unternehmens-, bzw. Abteilungsgrenzen hinaus. Ein technischer Ansatz wird von METTLER/RABER gegeben und beschreibt den unternehmensübergreifenden Einsatz eines zentralen DWH, auf das mehrere Beteiligte entlang der Wertschöpfungskette zugreifen und das sie für Analysen nutzen können. Die Synchronisation dieses DWH mit den lokalen Datenbeständen der Unternehmen findet dabei mit Hilfe bereits existierender Systeme zur Kooperation statt. Als Beispiel nennen die Autoren eine E-Procurement-Plattform.<sup>318</sup> Der wirtschaftliche Mehrwert ergibt sich dabei aus der Speicherung eigentlich flüchtiger Daten operativer Prozesse, die erst durch die gemeinsame Bereitstellung und Interpretation an Bedeutung gewinnen und Wissen über den Gesamtprozess schaffen sollen, was wiederum zu einer Erhöhung der Zuverlässigkeit reibungsloser Abläufe führen soll.<sup>319</sup>

Ebenfalls zur Unterstützung der Abläufe in Wertschöpfungsketten dient die von WERNER ET AL. vorgeschlagene Lösung eines Cockpits, das überbetriebliche Sichtweisen auf die Daten der Beteiligten ermöglicht. Als wirtschaftliche Motivation beschreiben die Autoren dabei die Entwicklung der Wertschöpfungsketten hin zu nachfrageorientierten Systemen und identifizieren CBI als notwendige Entwicklung zur unternehmensübergreifenden Kontrolle und Analyse von Daten und Prozessen.<sup>320</sup> Auf diese Lösung beziehen sich KLARMANN/ROTH/FRANCZYK,

---

<sup>316</sup> Vgl. Martins/daCosta/deSousa (2012): 103f.; sowie dort zitiert: Berthold et al. (2010).

<sup>317</sup> Auch MINUTO ESPIL/VAISMAN nennen Regierungsorganisationen als möglichen Einsatzbereich für multidimensionale Datenbetrachtung im Kontext verteilter OLAP-Systeme (vgl. Abschnitt 4.2.2).

<sup>318</sup> Vgl. Mettler/Raber (2011): 2f. Die beschriebene Plattform für die elektronisch gestützte Beschaffung ist myOpenFactory (<http://www.myopenfactory.com/>), eine kommerzielle Software, die den Standard *Electronic Data Interchange* (EDI) zur Kommunikation implementiert.

<sup>319</sup> Vgl. Mettler/Raber (2011): 1.

<sup>320</sup> Vgl. Werner et al. (2010): 1587.



die am Beispiel eines Kontraktdienstleisters<sup>321</sup> erläutern, wie eine über mehrere Unternehmen verteilte Entscheidungsfindung die Schaffung einer integrierten Sichtweise eines strategischen Netzwerks stärken kann.<sup>322</sup>

Die weiteste Öffnung des Analystenkreises wird von LIU/DANIELS beschrieben. Die Autoren betrachten die Möglichkeiten zur gemeinsamen Analyse in strategischen Unternehmensnetzwerken und argumentieren, dass erst die gemeinschaftliche Analyse mehrerer Unternehmen die Erkennung bestimmter Muster ermöglicht.<sup>323</sup> Als beispielhaften Anwendungsfall verweisen sie auf das *Fraud Information System Holland* (FISH).<sup>324</sup> Die Bereitschaft, nicht nur Teile der eigenen Daten zu teilen, sondern den Bestand auch für gemeinschaftliche Analysen zu nutzen, ermöglicht den Austausch von Erfahrungen im Analyseprozess und damit auch die Erkennung schwach ausgeprägter Muster.<sup>325</sup>

### 3.3.4 Anwendung auf unternehmensübergreifende Kooperation in Analyseprozessen

Von den beschriebenen Formen von CBI weisen nur CBI-PD und CBI-PA eine relevante Ausweitung des Teilnehmerkreises über Unternehmensgrenzen hinweg auf. Es zeigt sich, dass nur wenig Literatur existiert, die gezielt eine unternehmensübergreifende Kooperation im Bereich BI thematisiert, insbesondere mit Fokus auf die Diskussion einer Lösungsumsetzung.

Eine Betrachtung der Literatur mit Bezug auf die vier Sichten (TOFE) zeigt auch, dass für alle Sichten ein Forschungspotenzial besteht. So existiert nahezu keine empirische Aussage zur Höhe des (wirtschaftlichen) Nutzens der Systeme. Aus technischer Sicht wurden Ideen zur Netzgestaltung formuliert, eine (prototypische) Umsetzung bisher jedoch nicht oder nicht vollständig<sup>326</sup> durchgeführt. Die wesentliche fachliche Fragestellung ist die der Integration bzw. Überführung von Daten. Während in einem einzelnen DWH-System eine integrierende Schicht

<sup>321</sup> Aufgabe eines Kontraktdienstleisters sind „(...) die Gestaltung und der Betrieb eines firmenspezifischen Netzwerks und die Integration in die bestehende Prozesslandschaft des Auftraggebers.“ (Klarmann/Roth/Franczyk (2013): 28f.). Der Kontraktdienstleister bezieht die dafür notwendigen Ressourcen jedoch seinerseits wieder von extern, sodass er lediglich eine koordinierende Rolle wahrnimmt, für die eine integrierte Sichtweise auf die Daten und Kennzahlen aller Beteiligten zwingend notwendig ist (vgl. Klarmann/Roth/Franczyk (2013): 29, 32).

<sup>322</sup> Vgl. Klarmann/Roth/Franczyk (2013): 32–34.

<sup>323</sup> Vgl. Liu/Daniels (2012): 3.

<sup>324</sup> Vgl. Liu/Daniels (2012): 4; sowie die Erläuterungen in Fußnote 280.

<sup>325</sup> Vgl. Liu/Daniels (2012): 2f.

<sup>326</sup> Vgl. Berthold et al. (2010): 1.

existiert, die mehrere Datenquellen in eine zentrale Datenbank überführt, muss beim Zugriff auf mehrere (fremde) DWH-Systeme ein Mechanismus existieren, der ‚im Nachhinein‘ eine Einheitlichkeit des Zugriffs sicherstellt. Die Messung und Bewertung der Güte dieses Zugriffs sind dabei noch ungeklärt.<sup>327</sup> Organisatorische Fragestellungen ergeben sich zwangsläufig durch die Beteiligung mehrerer Partner an einem Netzwerk.

Die von mehreren Autoren betonte Relevanz einer unternehmensübergreifenden BI<sup>328</sup> stützt aber die Ausführungen zu unternehmensübergreifender Kooperation und die Beschreibung des Fallbeispiels in dieser Arbeit.<sup>329</sup> Für die gemeinschaftliche Auswertung von Daten mehrerer Unternehmen kann somit konzeptionell ein CBI-System herangezogen werden. Deutlich wird dabei, dass einer gemeinschaftlichen Analyse zunächst eine gemeinschaftliche Datenhaltung, zumindest im Sinne eines übergreifenden Zugriffs auf die Daten, vorausgehen muss.

Als Grundlage für die folgenden Überlegungen werden daher zunächst die Konzepte und Publikationen im Bereich des CBI-PD verwendet. Dabei stellt sich die beschriebene Situation als gemeinschaftlicher Zugriff auf unabhängig voneinander vorgehaltene multidimensionale Datenstrukturen dar, auf die auf der Präsentationsebene mit Hilfe von OLAP zugegriffen werden kann. Die für BI wesentliche Idee eines Data Warehouse wird dabei aufgegriffen und durch die beschriebenen kooperativen Mechanismen und Konzepte zur Gestaltung von Netzwerken ergänzt.

---

<sup>327</sup> Vgl. Rizzi (2012): 202.

<sup>328</sup> Vgl. Abschnitt 3.3.3, dabei insbesondere Abschnitt 3.3.3.2.

<sup>329</sup> Vgl. Abschnitt 2.4. Ergänzend sei auf das Projekt ‚CEUS‘ (für: ‚Computerbasiertes Entscheidungsunterstützungssystem für die Hochschulen in Bayern‘) verwiesen, in dem verschiedene bayerische Hochschulen lokale DWHs nutzen, um diese regelmäßig in ein zentrales DWH zu überführen und Informationen über Lehre und Forschung auszutauschen. Es liegt dabei eine verteilte DWH-Architektur vor (vgl. zu Architekturformen Abschnitt 4.2.1). Die zentralen, formulierten Anforderungen entsprechen den bisher beschriebenen Anforderungen an unternehmensübergreifende BI-Netzwerke: Unabhängigkeit, Flexibilität, (zeitliche) Entkopplung, Verwendung vertrauensbasierter Schnittstellen und sukzessive Erweiterung (vgl. Sinz/Ulbrich-vom-Ende (2015)).

## 4 Konzept eines unternehmensübergreifenden Business-Intelligence-Netzwerks

Wie in der Ergebnisbetrachtung der vorhergehenden Literaturrecherche formuliert, besteht ein starkes Ungleichgewicht in der Anzahl der Ausarbeitungen zu kooperativen Analyseprozessen in bzw. zwischen Unternehmen. Umfangreiche Forschung im Bereich unternehmensübergreifender BI liegt dabei nicht vor. Dieses Erkenntnis wird auch in den thematisch relevanten Arbeiten formuliert. GOLFARELLI ET AL. beschreiben ein Fehlen von Publikationen im Bereich der Integration und unternehmensübergreifenden Verwendung von DWH-Lösungen. Sie verweisen darauf, dass, falls die Fragestellung einer einheitlichen Datenbetrachtung mehrerer Parteien aufkommt, vorwiegend komplexe ETL-Prozesse zur Heterogenitätsminderung eingesetzt werden.<sup>330</sup> BERGAMASCHI ET AL. geben als relevante Autoren in Bezug auf die Integration unterschiedlicher DWH-Systeme BANEK ET AL., CABIBBO/TORLONE und GOLFARELLI ET AL. an und verweisen ansonsten auf einen Mangel an Forschung im Bereich unternehmensübergreifender BI.<sup>331</sup> In den genannten Publikationen werden nur einzelne – teilweise aber mehrere – Aspekte der Thematik herausgegriffen. Als vergleichsweise weit ausgearbeitet ist dabei die Arbeit von RIZZI zu dem Konzept des BIN<sup>332</sup> hervorzuheben, auf die sich GOLFARELLI ET AL. beziehen.

Im vorliegenden Kapitel wird zunächst eine begriffliche Schärfung unternehmensübergreifender BI vorgenommen. Auf Basis der genannten Publikationen sowie der Ergebnisse der Literaturrecherche zu CBI wurde eine weiterführende kombinierte Vorwärts-/ Rückwärts-Suche nach thematisch verwandten Publikationen durchgeführt, um Arbeiten zu identifizieren, die einzelne Aspekte des Themas behandeln und als Grundlage eines umfassenderen Konzepts für unternehmensübergreifende BI dienen können.<sup>333</sup> Die durchgeführte Suche hat dabei keinen weiteren Ansatz hervorgebracht, der die Thematik umfassend behandelt, mit *Exploratory OLAP* aber ein Konzept aufgezeigt, das ähnliche Ideen und Probleme

---

<sup>330</sup> Vgl. Golfarelli et al. (2012a): 153f.

<sup>331</sup> Vgl. Bergamaschi et al. (2011): 38f. Die genannten Quellen sind: Banek et al. (2007), Cabibbo/Torlone (2004), Golfarelli et al. (2012b) und Torlone (2008).

<sup>332</sup> Zu dem Konzept des BIN vgl. Rizzi (2012) sowie Abschnitt 3.3.3.2 und die ausführliche Darstellung in Abschnitt 4.1.

<sup>333</sup> Das Vorgehen folgt dabei den Empfehlungen von WEBSTER/WATSON zur Erstellung von Reviews. Vgl. dazu allgemein die Ausführungen in Abschnitt 3.3.1.1.

auf Gesamtkonzeptebene aufbringt. Es schließt sich daher an die Begriffsdefinition ein Überblick abgeleiteter Fragestellungen aus den Konzepten des BIN und des Exploratory OLAP an. Im weiteren Verlauf des Kapitels werden relevante Ansätze, die entweder Teilaspekte oder verwandte Themen behandeln, vorgestellt und kritisch bewertet. Abschließend erfolgt die Entwicklung eines Gesamtmodells, das für den weiteren Verlauf der Arbeit einen Bezugspunkt verwendeter Methoden bietet.

#### 4.1 Cross-Company Business Intelligence

Es lässt sich erkennen, dass der Begriff CBI von populär-wissenschaftlichen Fachmedien mit großer Mehrheit für unternehmensinterne Entwicklungen verwendet wird – so wie es auch in den wissenschaftlichen Publikationen der Fall ist.<sup>334</sup> Es erscheint daher angemessen, für den weiteren Verlauf der Arbeit den Fokus auf unternehmensübergreifende Prozesse und Systeme auch sprachlich stärker herauszustellen. Da die Überlegungen dieser Arbeit zur gemeinschaftlichen Datenhaltung und -auswertung in Konsequenz auch gemeinsame Analysen unterstützen sollen, wird für die folgenden Ausführungen von ‚unternehmensübergreifender Business Intelligence‘ ohne weitere Einschränkung gesprochen. Um eine sprachliche Einheitlichkeit der in dieser Arbeit verwendeten Begriffe herzustellen,<sup>335</sup> findet die englische Entsprechung *Cross-Company Business Intelligence* (CCBI) Anwendung.<sup>336</sup> Entsprechend den Erkenntnissen aus Abschnitt 3.3 existieren in der Literatur bisher weder ein anderer dedizierter Terminus, noch ein entsprechendes, explizit umrissenes Forschungsfeld.<sup>337</sup> Für CCBI wird im Rahmen dieser Arbeit daher folgende Definition verwendet:

---

<sup>334</sup> Vgl. Müller (2013); Rohrmoser (2014): 28; Weber/Kolb (2014): 20; sowie die Ergebnisse der Literaturrecherche in Abschnitt 3.3.

<sup>335</sup> Die Problematik der Übersetzung einzelner Fachtermini aus dem Englischen und der entsprechende Umgang damit wurde bereits in Abschnitt 3.1 thematisiert.

<sup>336</sup> Analog zu der ‚eingedeutschten‘ Schreibweise von bspw. DSS wird auch hier der jeweils erste Buchstabe jedes Wortes groß geschrieben.

<sup>337</sup> Der Begriff ‚unternehmensübergreifende BI‘ wird zwar von Kemper/Baars/Mehanna verwendet, umschreibt dort jedoch einen praxisbezogenen Anwendungsfall und wird nicht als feststehender Ausdruck beschrieben oder im Weiteren angewandt (vgl. Kemper/Baars/Mehanna (2010): 232).

**Definition (6)** *Cross-Company Business Intelligence* umfasst Konzepte, Methoden und Systeme zur Unterstützung unternehmensübergreifender Datenhaltung, -bereitstellung, -präsentation und -analyse unter Berücksichtigung der Unabhängigkeit der Partner.

Die bisherige Definition (5) für CBI wird aufrechterhalten, da sie im Kontext des erarbeiteten Verständnisses in der Literatur weiterhin Gültigkeit behält. Es kann somit argumentiert werden, dass CCBI eine Untermenge der Ansätze von CBI beinhaltet. Eine weitere, ausführliche Abgrenzung bzw. Betrachtung der genauen Inhalte kann aufgrund fehlender Verwendung in der Literatur nur definitorisch vorgenommen werden. Es sei als Ansatzpunkt für eine solche Unterscheidung darauf hingewiesen, dass die Datenerfassung/-gewinnung explizit nicht Bestandteil der verwendeten Definition von CCBI ist, da die gemeinschaftlichen, unternehmensübergreifenden Prozesse nicht direkt Datenquellen Dritter auslesen (wie z. B. Open-Data-Quellen<sup>338</sup>), sondern diese nur mittelbar durch Systeme einzelner beteiligter Unternehmen eingebracht werden.

Der CCBI-Definition am nächsten kommt aus der identifizierten Literatur das BIN-Konzept, dessen grundlegende Beschreibung im Wesentlichen in den drei Arbeiten GOLFARELLI ET AL. (2011), GOLFARELLI ET AL. (2012B) und RIZZI (2012) vorgenommen wird. Die Autoren stellen dabei ein Netzwerk aus autonomen Netzwerkteilnehmern vor, die BI-Funktionalitäten über ein P2P-Netzwerk<sup>339</sup> austauschen. Im Fokus steht dabei das Senden und Empfangen von OLAP-Abfragen über das Netzwerk verbundener DWH-Systeme anderer Peers. Die Autoren richten daher die Strukturbeschreibung des Netzwerks und der Teilnehmer auf den Datenfluss einer OLAP-Abfrage (sowohl intern als auch extern aus). Die Abfrage wird von einem Nutzer generiert und dann intern an das eigene DWH geleitet sowie an die Partner im Netzwerk weitergegeben (vgl. Abb. 26). Es ist dabei nicht

---

<sup>338</sup> Vgl. Abschnitt 3.3.2.

<sup>339</sup> Ein P2P-Netzwerk ist die Verknüpfung mehrerer, gleichgestellter Teilnehmer. Es genügt den drei Eigenschaften (1) gleichzeitiger Server- und Clientfunktionalität aller Beteiligten (2) einer fehlenden zentralen Kontrollinstanz sowie (3) vollständiger Selbstbestimmtheit der jeweils eigenen Aktivitäten. Verbreitung hat dieser Netzwerktypus unter anderem mit Filesharing („Dateiaustausch“)-Diensten um die Jahrtausendwende gefunden (vgl. Miller (2001); Schoder/Fischbach (2002): 587–589; sowie ergänzend Barkai (2001) zit. nach Schoder/Fischbach (2002): 587). Eine ausführlichere Beschreibung erfolgt im Rahmen der Referenzmodellerstellung in Abschnitt 4.5.1. Es sei darauf hingewiesen, dass hier – wie auch bei SCHODER/FISCHBACH – *unstrukturierte* P2P-Netzwerke gemeint sind. *Strukturierte* P2P-Netzwerke weisen vordefinierte Verbindungsstrukturen auf und genügen damit den Eigenschaften (2) und (3) nicht oder nur teilweise.

gefordert, dass alle Parteien direkt miteinander verbunden sind, sodass auch Abfragewege mit ‚Zwischenstationen‘ berücksichtigt werden.

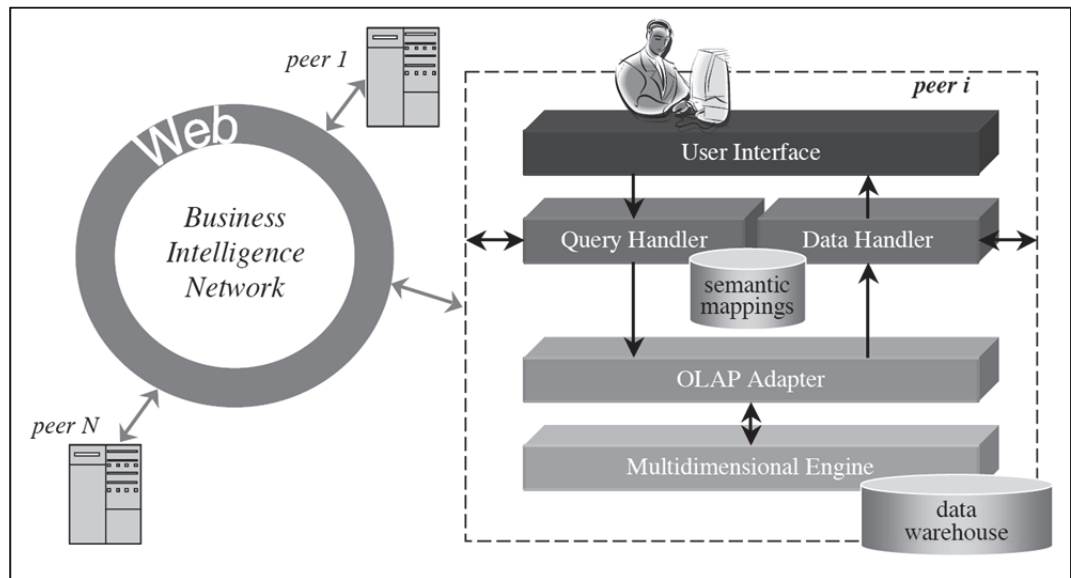


Abb. 26: BIN-Architektur

(Quelle: Rizzi (2012): 194)

Die Autoren fokussieren ihre Arbeit auf das Finden einer effektiven ‚query reformulation‘, d. h. einer Umformulierung der Abfrage von einem System zum nächsten. Sie begründen einen Großteil ihrer Arbeit in den Ideen zum Zugriff der Teilnehmer von *Peer-Data-Management-Systemen* (PDMS).<sup>340</sup> Dabei wird eine unterliegende relationale Datenstruktur angenommen.<sup>341</sup> Neben der technisch geprägten Ausarbeitung formulieren die Autoren weitere, allgemeine Fragestellungen und Probleme eines BIN:<sup>342</sup>

<sup>340</sup> Der Begriff ‚peer data management system‘ basiert auf einer Arbeit von HALEVY ET AL. und beschreibt die Verknüpfung unterschiedlicher Datenbanken (vgl. Halevy et al. (2004): 787f.). Im Kontext von PDMS ist in den letzten Jahren kontinuierlich Forschung betrieben worden, sodass dort auch entsprechende Hinweise zu den Fragestellungen im Bereich BI-Netzwerke thematisiert werden. Einzelne Aspekte werden in den entsprechenden Abschnitten dieser Arbeit aufgenommen (vgl. Abschnitt 4.3 sowie insbesondere Abschnitte 5.2–5.4). Da PDMS aber ein zunächst rein relational geprägtes Umfeld betrachten, lassen sich die Arbeiten nicht direkt auf den BI-Kontext übertragen (vgl. Golfarelli et al. (2011): 246–249), selbst dann nicht, wenn BI-relevante Komponenten wie Aggregat-Abfragen – d. h. Abfragen von Hierarchieknoten – thematisiert werden, da diese zunächst rein technische Fragestellungen betrachten, nicht aber die semantische Implikation einer multidimensionalen Auswertung berücksichtigen (vgl. z. B. die erste Arbeit zu diesem Thema: Cohen/Nutt/Sagiv (2006)).

<sup>341</sup> Vgl. hierzu die Hinweise zu ROLAP in Abschnitt 3.1.3.

<sup>342</sup> Vgl. hierzu und zur folgenden Aufzählung Golfarelli et al. (2011): 246f.

- Die Abfragebearbeitung sollte einer Geschwindigkeitsoptimierung unterzogen werden, um grundlegende Anforderungen an OLAP-Systeme weiter erfüllen zu können.<sup>343</sup>
- Die Resultate der Abfragen müssen zusammengeführt werden. Dabei muss eine Sicherstellung der Interpretierbarkeit berücksichtigt werden, insbesondere bei heterogenen multidimensionalen Schemata.
- Die Nutzbarkeit der zurückgelieferten Daten sollte messbar gemacht werden.
- Neben der Nutzbarkeit sollte auch die Qualität der Daten<sup>344</sup> nachvollziehbar dargestellt werden, ggf. mit einer Rückverfolgung über den gesamten Anlieferungsweg im Netzwerk.
- Es sind Sicherheitskonzepte, sowohl technischer als auch organisatorischer Art, für die Datenübertragung und den Zugriff auf die Daten zu entwickeln und zu implementieren.

Eine ähnliche Übersicht an Fragestellungen findet sich in den Ausführungen von ABELLÓ ET AL. zu Exploratory OLAP. Diese Bezeichnung findet sich üblicherweise in der Literatur nicht als feststehender Begriff, sondern als Beschreibung einer regulären Nutzung von OLAP-Funktionalitäten, ggf. in Kombination mit weiteren explorativen Analysemethoden.<sup>345</sup> Bei den genannten Autoren steht er jedoch für den Einsatz von OLAP-Werkzeugen, die neben unternehmensinternen Daten auch auf bisher nicht verwendete, externe Datenquellen zugreifen. Dabei verweisen ABELLÓ ET AL. auf zumeist öffentlich zugängliche, zumindest aber aus einem nicht-zielgerichteten (öffentlichen) Datenbestand generierte Daten, wie sie bspw. von Open-Data-Initiativen oder sozialen Netzwerken vorgehalten und bereitgestellt werden.<sup>346</sup> Die Autoren beschreiben im Weiteren die Entwicklung eines um semantische Funktionen angereicherten ETL-Prozesses zur Integration dieser Daten in den mit OLAP zu durchsuchenden Datenbestand. Eine auf diesen Ideen basierende Fallstudie wurde von IBRAGIMOV ET AL. erarbeitet und verwen-

---

<sup>343</sup> Vgl. Abschnitt 3.1.3.

<sup>344</sup> Zum Begriff der Qualität/Datenqualität vgl. Abschnitt 5.1.

<sup>345</sup> Vgl. Levera/Barán/Grossman (2004): 457; Priebe/Pernul (2000): 1f.

<sup>346</sup> Vgl. Abelló et al. (2014): 1.



det eine mit offenen Austauschformaten<sup>347</sup> spezifizierte Datenquelle als Basis zur temporären Generierung von OLAP-Würfeln.<sup>348</sup> In beiden Publikationen wird somit zwar kein Austausch zwischen Unternehmensquellen beschrieben, dennoch ergeben sich eine Reihe von Fragestellungen und identifizierten Problemen, die auch bei der Verknüpfung bereits strukturierter Unternehmens-DWH-Systeme und damit im Bereich CCBI relevant sind, wobei die Autoren teilweise bereits direkten Bezug auf eine Anwendung im BI-Bereich nehmen.<sup>349</sup>

- Zwischen den unterschiedlichen Ontologien der Systeme, die ggf. erst noch gestaltet werden müssen, sind Mappings, d. h. Zuordnungs- und Überleitungsregeln zu definieren. Eine automatische Unterstützung ist eine noch weitergehende Forderung.
- Die bisherigen Ausführungen beziehen sich auf das Semantic Web, ein mit semantischen Informationen angereichertes WWW. Hierarchien, wie sie in einem DWH und in OLAP-Analysen üblich sind, sind dort nicht vorgesehen.
- Bisherige Konzepte unterstützen keinen Zeitbezug der Daten. Das Problem veränderlicher Dimensionen (und allgemein Schemata) ist aber eine der zentralen Herausforderungen im Bereich multidimensionaler Datenhaltung.<sup>350</sup>
- Aus den definierten Strukturen zum Zugriff auf externe Daten lassen sich derzeit noch keine passenden ETL-Prozesse ableiten, die einen dauerhaften Zugriff gewährleisten.
- Unberücksichtigt bleibt – wie auch bei den BINs – ein Umgang mit nicht-relationalen Datenquellen.
- Es kann auf Grund der unabhängigen Datenquellen kein globales Datenschema implementiert werden. Ein Umgang mit autonomen, lokalen Schemata ist aber schwieriger und weniger komfortabel.<sup>351</sup>

---

<sup>347</sup> Verwendet wird das Resource Description Framework (RDF) als mittlerweile offiziell vom World Wide Web Consortium (W3C) vorgeschlagenes Austauschformat für Daten auf höheren semantischen Ebenen. Das W3C ist die international anerkannte Standardisierungsorganisation für das World Wide Web (WWW) als Teil des Internets (vgl. W3C (2014); W3C (2015)).

<sup>348</sup> Vgl. Ibragimov et al. (2014): 2–7.

<sup>349</sup> Vgl. für die folgende Aufzählung Abelló et al. (2014): 17–19.

<sup>350</sup> Vgl. Bauer/Günzel (2013): 260–265.

<sup>351</sup> Vgl. hierzu auch die Ausführungen zu ‚local-as-view‘- und ‚global-as-view‘-Ansätzen in Abschnitt 4.3.2.



- Ein automatisiertes Schlussfolgern („reasoning“) in den Daten ist daher nur in lokalen Instanzen möglich.

Als letzten Punkt, der im Rahmen einer Betrachtung des gesamten Ansatzes von besonderer Bedeutung ist, stellen die Autoren die Frage, ob der Ansatz, Techniken des Semantic Web einzusetzen, um DSS mit Daten zu beliefern, überhaupt ein sinnvoller Weg ist. Sie kommen dabei zu dem Schluss, dass es möglicherweise sinnvoller ist, einen Teil der Autonomie (der Quellen) aufzugeben, bzw. diese Aufgabe einzufordern, und eine durch eine zentrale Instanz festgelegte Grundstruktur zu verwenden.<sup>352</sup>

Wie gezeigt, beziehen sich nahezu sämtliche Ausführungen zu CBI und CCBI explizit oder implizit auf ein OLAP-basiertes oder -ähnliches Modell und stärker auf anwenderorientierte und analysefokussierte Fragestellungen als auf Fragen der eigentlichen Datenspeicherung.<sup>353</sup> Für die folgenden Betrachtungen wird dieser Fokus aufrechterhalten, da die Datenhaltung im CCBI-Kontext nicht zum Selbstzweck (z. B. zur Erfüllung gesetzlicher Speicherpflichten), sondern als reine Grundlage für eine Analyse durchgeführt wird. Eine Übersicht entsprechender Ansätze und ein Ausblick auf komplexere Modelle wird im folgenden Abschnitt gegeben. Die nachfolgenden Abschnitte gehen auf weitere, thematisch geordnete Konzepte und Lösungen ein, die zur Entwicklung eines CCBI-Konzepts der -Systems relevante Aspekte beschreiben und dabei auch die in beiden vorgestellten Ansätzen (BIN und Exploratory OLAP) aufgebrachten Fragestellungen in Teilen adressieren.

## 4.2 Paralleles und verteiltes Arbeiten im Data-Warehouse-Kontext

Konzepte paralleler und verteilter Arbeit finden sich im Kontext von IT zunächst auf Ebene einzelner Recheneinheiten und Speichersysteme. Die mangelnde Leistungsfähigkeit einzelner Computersysteme wird dabei kompensiert durch den simultanen Einsatz mehrerer Systeme oder Komponenten. Paralleles Rechnen (engl.: „parallel computing“) bezeichnet daher zunächst nur die gleichzeitige Nutzung mehrerer Recheneinheiten zur Bewältigung einer Aufgabe.<sup>354</sup> Zur Implementierung dieses Vorgehens lassen sich drei grundlegende Architekturmodelle der

---

<sup>352</sup> Vgl. Abelló et al. (2014): 19.

<sup>353</sup> Vgl. Abschnitt 3.3.

<sup>354</sup> Vgl. Eckstein (2013): 1103f.; Malek (2006): 199f.

beteiligten Computersysteme unterscheiden. Abb. 27 stellt dar, wie Recheneinheiten („cpu“), Verarbeitungsspeicher („memory“/„mem“) und dauerhafte Datenhaltung („disk“) miteinander gekoppelt werden können. Eine Shared-Memory-Architektur verfolgt den Ansatz, ein integriertes System durch den verstärkten Einsatz miteinander verbundener Komponenten zu bilden und zu beschleunigen. Der gegenteilige Ansatz findet sich bei einer Shared-Nothing-Architektur. Hier werden vollständig unabhängige Systeme gebildet, die über ein Netzwerk verbunden werden. Als ‚Zwischenlösung‘ kann eine Shared-Disk-Architektur verstanden werden, bei der unabhängige Rechen-/Speichereinheiten auf einen gemeinsamen Datenbestand zugreifen.<sup>355</sup>

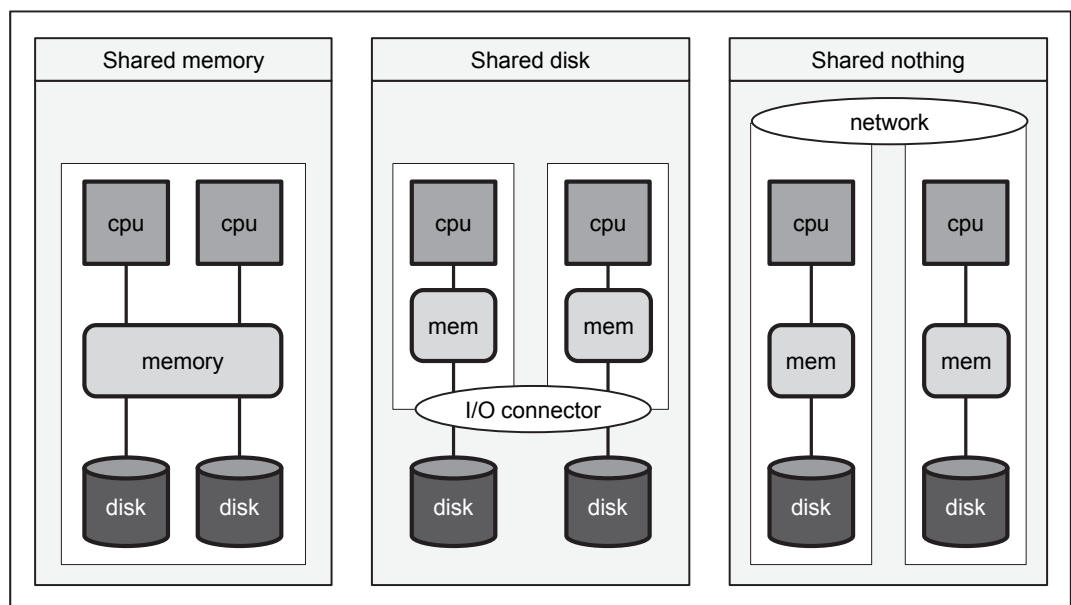


Abb. 27: Unterschiedliche Architekturen für paralleles Rechnen  
(Quelle: In Anlehnung an Furtado (2009): 58f.)

Die Auswahl der Architektur hängt dabei von den gegebenen Rahmenbedingungen und von Geschwindigkeitsüberlegungen ab, da der langsamste Teil des Verbundes die Gesamtgeschwindigkeit maßgeblich beeinflusst.<sup>356</sup> Der Einsatz einer Shared-Nothing-Architektur wird dabei bereits seit den späten 80er-Jahren trotz der zu dieser Zeit geringen Netzwerkbandbreiten als nicht nachteilig betrachtet.<sup>357</sup>

Eine einheitliche Abgrenzung zwischen ‚parallelen‘ und ‚verteilten‘ Ansätzen ist, trotz der Verwendung beider Begriffe, nicht immer gegeben und wird teilweise

<sup>355</sup> Vgl. Furtado (2009): 58–60.

<sup>356</sup> Vgl. Furtado (2009): 60.

<sup>357</sup> Vgl. Stonebraker (1986): 5, 9; Dewitt/Gray (1992): 97.

auch nicht mehr vorgenommen.<sup>358</sup> Zur Übertragung der Konzepte auf den Kontext von DWH-Systemen verwendet ein Großteil der Autoren jedoch die Bezeichnung ‚verteilte Systeme‘<sup>359</sup>, sodass im Weiteren auch von diesen gesprochen wird. Die folgenden Abschnitte stellen die entsprechenden Ansätze zu DWH-Systemen vor und zeigen mit P2P-basiertem OLAP eine explizit als Shared-Nothing-Architektur vorgesehene Struktur zur Abfrage von Daten auf.

#### 4.2.1 Verteilte Data-Warehouse-Systeme

Der in Abschnitt 3.1.2 dieser Arbeit vorgestellte Ansatz eines DWH-Systems stellt, wie dort beschrieben, nur eine mögliche, stark integrierte Form dar. Das zentrale DWH ist dabei die (einzige) Datenbasis für die Instrumente höher gelegener Schichten zur Analyse. Ähnlich den unterschiedlichen Architekturen für paralleles Rechnen lassen sich aber auch für ein DWH unterschiedliche Ansätze identifizieren. Insbesondere dann, wenn ein Unternehmen (noch) keine zentrale Datenbasis schaffen kann oder will, oder wenn Performance-Gründe gegen eine einzige, zentrale Datenhaltung sprechen, bietet sich das bewusste Schaffen verteilter Strukturen an. Der Gedanke eines DWH, das aus einzelnen, physisch verteilten Data Marts gebildet wird, die alle einem einheitlichen Schema folgen und auch in Summe angesprochen werden können, ist dabei bereits zu Beginn der systematischen DWH-Forschung aufgebracht worden.<sup>360</sup> Abb. 28 zeigt zum besseren Verständnis daher zunächst die generellen, unterschiedlichen Architektur-Konzepte im Rahmen von DWH-Systemen.

---

<sup>358</sup> JUAN ET AL. sprechen z. B. von „Parallel and Distributed Computing Systems (PDCS)“ (Juan et al. (2013): 47). Eine klare Unterscheidung der Begriffe nimmt MALEK vor, der bei verteilten Systemen grundsätzlich Unabhängigkeit voraussetzt (vgl. Malek (2006): 209).

<sup>359</sup> Üblicherweise wird der englische Begriff ‚distributed‘ verwendet (vgl. z. B. Akinde et al. (2003), Albrecht/Lehner (1998), Doka/Tsoumakos/Koziris (2011)).

<sup>360</sup> Vgl. Albrecht/Lehner (1998): 79.

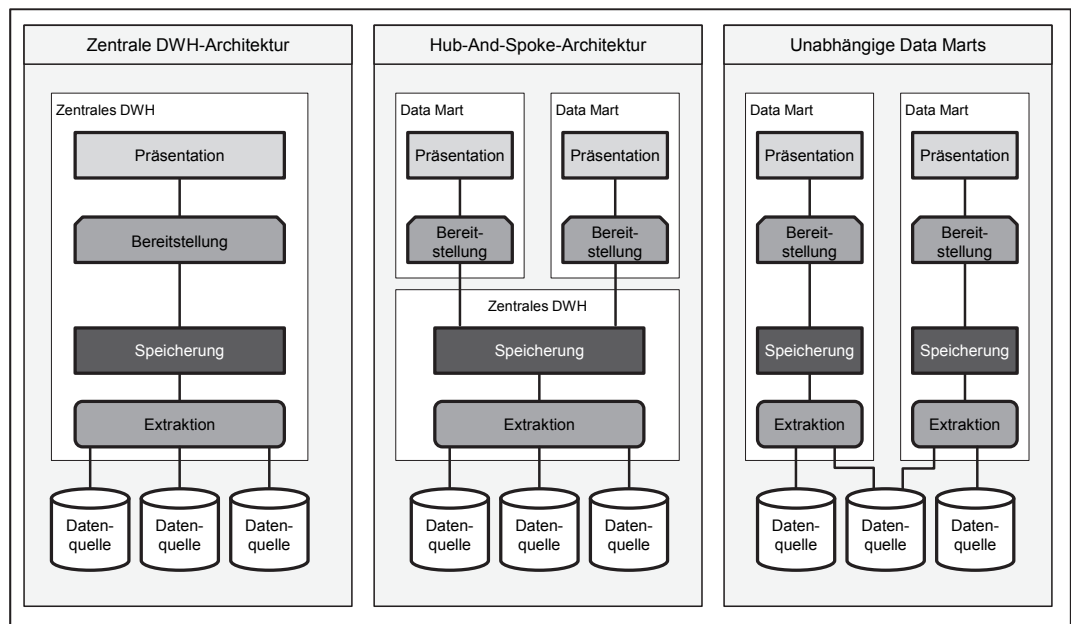


Abb. 28: Unterschiedliche DWH-Architekturen (1)

(Quelle: In Anlehnung an Sinz/Ulbrich-vom-Ende (2010): 188–190)

Es lässt sich bei der Betrachtung eine Nähe zu den vorgestellten Konzepten parallelen Rechnens erkennen. Die zentrale DWH-Architektur stellt eine vergleichsweise monolithische Struktur dar, die durch eine Hub-and-Spoke-Architektur aufgebrochen wird, da dort nur die unteren Schichten bis einschließlich der Datenhaltung zentral behandelt werden und die Datenbasis im Anschluss mehrere Data Marts mit Daten beschickt. Vollständig autonom agieren unabhängige Data Marts, die auch den ETL-Prozess separat abhandeln. Eine einheitliche Sichtweise auf die Daten entsteht bei der letzten Form aber erst mit der Hinzunahme einer Data-Mart-Bus-Architektur, die die einzelnen Systeme aneinander koppelt (vgl. Abb. 29). Um daraus einen einheitlich zugreifbaren Datenbestand zu formen, muss eine föderierte DWH-Architektur geschaffen werden, die die einzelnen Data Marts virtuell konsolidiert.<sup>361</sup>

<sup>361</sup> Vgl. Sinz/Ulbrich-vom-Ende (2010): 187–193. Die Autoren stellen neben den genannten Architekturen noch ein ‚hierarchisches DWH-System‘ vor, das das föderierte Konzept um eine zentrale integrierte Datenhaltungsschicht ergänzt. Dies ist für die weiteren Ausführungen nicht relevant. Für Vor- und Nachteile der einzelnen Ansätze sowie weiterführende Literatur vgl. die angegebene Quelle.

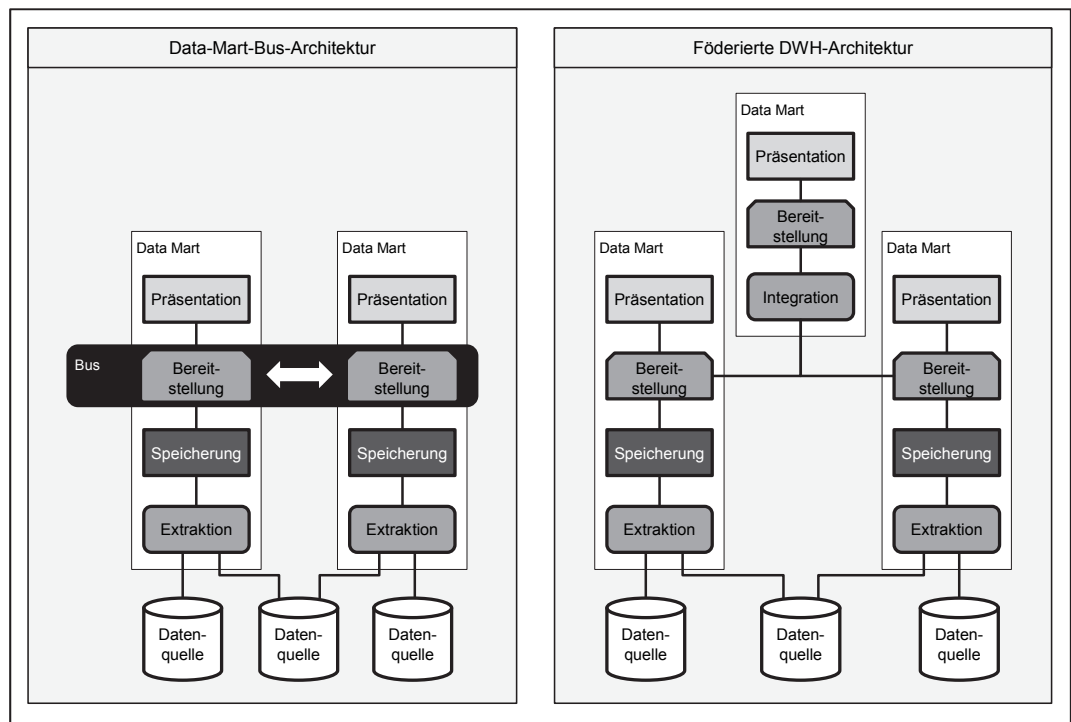


Abb. 29: Unterschiedliche DWH-Architekturen (2)

(Quelle: In Anlehnung an Sinz/Ulbrich-vom-Ende (2010): 192f.)

FURTADO zeigt in einer Übersicht zu parallelen und verteilten DWH-Systemen, dass unter diesen Begriffen sowohl Ansätze zur effizienten Verteilung von Datenbanken zentraler DWH-Systeme als auch zur Nutzung bereits physisch getrennt existierender Systeme – entsprechend einer der Data-Mart-Architekturen – diskutiert werden.<sup>362</sup> Für die Betrachtung unternehmensübergreifender Systeme werden daher Arbeiten zu beiden Themengebieten vorgestellt. Abzugrenzen sind hingegen Arbeiten zu verteilten Datenbank- oder Dateisystemen, die in den letzten Jahren vermehrt mit der unter dem Begriff *Big Data* subsummierten Forschung zu großen, heterogenen und oftmals verteilt gespeicherten Datenmengen veröffentlicht wurden.<sup>363</sup> Obwohl Big Data, insbesondere im Bereich der Analyse, als BI-nahe Thematik verstanden werden kann, liegt dort der Fokus auf unstrukturierten,

<sup>362</sup> Vgl. Furtado (2009): 71–75.

<sup>363</sup> Eine Übersicht zum aktuellen Stand der Forschung und zu verschiedenen Verständnissen liefern CHEN/MAO/LIU (vgl. Chen/Mao/Liu (2014)). Unter Annahme der – auf der ursprünglichen 3V-Definition aufbauenden (vgl. zur Übersicht und Einführung Klein/Tran-Gia/Hartmann (2013): 319–321) – Definition von GANTZ/REINSEL (vgl. Gantz/Reinsel (2011): 6), dass Big Data durch (1) große Datenmengen („volume“) in (2) unstrukturierter Form („variety“) die (3) schnell veränderlich sind („velocity“) und (4) wenig Wert in unkonsolidierter Form („value“) aufweisen, beschrieben wird, können diese sowohl als Datenquelle für DWH-Systeme auftreten als auch durch die typischerweise im BI-Bereich eingesetzten Analysemethoden verarbeitet werden. Dabei besteht aber immer nur eine Korrespondenz von DWH-Konzepten und Big-Data-Fragestellungen, keine direkte Zusammenführung (vgl. Chen/Mao/Liu (2014): 172f., 183, 192–195).

semantisch oft unklaren Datenbeständen,<sup>364</sup> was dem Gedanken des DWH, so wie in dieser Arbeit formuliert und verstanden, widerspricht.

Eine allgemeine Betrachtung verteilter DWH-Systeme nehmen ALBRECHT/LEHNER vor. Sie beschreiben ein DWH, das aus unterschiedlichen, an mehreren globalen Standorten verteilten Data Marts besteht und über eine OLAP-Middleware<sup>365</sup> angesprochen wird. Dabei setzen sie voraus, dass ein einheitliches Schema in allen Data Marts vorherrscht. Zur Verringerung der Netztransaktionskosten schlagen sie eine teilredundante Datenhaltung vor. Sie unterscheiden dabei zwischen vertikaler Redundanz, d. h. der Vorberechnung und Speicherung einer Anzahl an Aggregaten, und horizontaler Redundanz, d. h. der Replikation errechneter Aggregate auf den Daten, sodass Abfragen zumindest zum Teil ohne Zugriff über das Netzwerk beantwortet werden können. Als offene Fragestellung wird die Auswahl der zu errechnenden Aggregate genannt.<sup>366</sup>

Eine redundante Datenhaltung wird auch von FURTADO thematisiert, der von einer relationalen Speicherung der DWH-Tabellen<sup>367</sup> ausgeht und mit der erheblichen Diskrepanz des Speicherbedarfs von Faktentabellen (sehr groß) zu Dimensionstabellen (klein) argumentiert. Er schlägt vor, die Dimensionstabellen in allen beteiligten Systemen redundant vorzuhalten, die Faktentabelle jedoch zu partitionieren, d. h. in disjunkte Mengen zu splitten und auf die Systeme zu verteilen. Dabei kommt sowohl eine horizontale als auch vertikale Partitionierung in Betracht. Im erstgenannten Fall erhalten alle Systeme eine bestimmte Anzahl vollständiger Datensätze, im zweiten Fall wird nach Spalten geteilt, was allerdings notwendigerweise zu einer redundanten Haltung der Primärschlüssel führt.<sup>368</sup> Neben einer Workload-basierten Partitionierung, die ein paralleles Verarbeiten von Abfragen

---

<sup>364</sup> Vgl. Chen/Chiang/Storey (2012): 1166–1168.

<sup>365</sup> Middleware bezeichnet nach TRESCH zumeist „eine mehr oder weniger hoch entwickelte Kommunikationskomponente, die Clients mit Servern verbindet“ (Tresch (1996): 250), aber auch komplexere Aufgaben wahrnehmen kann (vgl. ebd.). Der Begriff hat mit der weiten Verbreitung von Client-Server-Architekturen in den 1990er-Jahren größere Verwendung gefunden und kann in einer DWH-Umgebung z. B. auf Ebene der Datenbereitstellung zur Kommunikationsunterstützung mit separaten Client-OLAP-Anwendungen dienen.

<sup>366</sup> Vgl. Albrecht/Lehner (1998): 79–81, 84f.

<sup>367</sup> Vgl. zur relationalen Speicherung Abschnitt 3.1.3.

<sup>368</sup> Vgl. Furtado (2009): 63–65.

ermöglicht – sofern die Inter-Systemkommunikation schnell genug ist – können auch andere Strategien zum Einsatz kommen, z. B. nach Dimensionselementen.<sup>369</sup>

Die Annahme eines nicht durch die Verbindung der einzelnen Bauteile in seiner Geschwindigkeit limitierten Systems treffen auch WEHRLE/MIQUEL/TCHOUNIKINE. Sie verwenden das etablierte Konzept eines *Computing Grid* (CG) zur Verteilung der DWH-Daten und zur Unterstützung einer Vielzahl gleichzeitiger Zugriffe und greifen damit das Problem großer, dezentral organisierter Datenmengen auf, für die existierende IT-Infrastrukturen nicht ausgelegt sind.<sup>370</sup> Die verwendete Definition des CG als Hard- und Software, die verlässlich, dezentral und auf offenen Protokollen basierend, zugreifbar ist,<sup>371</sup> ähnelt dabei stark dem aktuellen Verständnis von *Cloud Computing*, das im Bereich flexibler IT-Infrastrukturen Verwendung erfährt.<sup>372</sup> Vorteile sind hierbei sowohl allgemein als auch für DWH-Systeme im Geschwindigkeits- wie im Kostenaspekt zu erzielen.<sup>373</sup> Die Autoren verweisen explizit darauf, dass sie die Verteilung einer zentralen DWH-Lösung thematisieren, das Vorgehen aber auch auf bereits verteilte, nicht rekonfigurierbare Systeme anwendbar ist.<sup>374,375</sup>

---

<sup>369</sup> Denkbar ist z. B., dass bei mehreren Standorten weltweit die Daten mit der Ausprägung ‚Deutschland‘ (in einer angenommenen Geografie-Dimension) auf dem deutschen System gespeichert werden, in der Annahme, dass dort eine hohe lokale Abfragelast erzeugt wird.

<sup>370</sup> Vgl. Wehrle/Miquel/Tchounikine (2005): 203f.

<sup>371</sup> Vgl. Foster (2002); Foster/Kesselman (2004) zit. nach Wehrle/Miquel/Tchounikine (2005): 204.

<sup>372</sup> Eine einheitliche Definition zu Cloud Computing ist nicht gegeben. International verbreitet und auch vom Bundesamt für Sicherheit in der Informationstechnik mit aufgenommen ist die Definition des amerikanischen National Institute of Standards and Technology: „Cloud computing is a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction. This cloud model is composed of five essential characteristics, three service models, and four deployment models.“ (Mell/Grance (2011): 2; vgl. ergänzend Bundesamt für Sicherheit in der Informationstechnik (2012): 14–16).

<sup>373</sup> Vgl. Shah (2014): 118.

<sup>374</sup> Vgl. Wehrle/Miquel/Tchounikine (2005): 209.

<sup>375</sup> PARK/JOO stellen ebenfalls eine auf Cloud Computing basierende Architektur vor, bei der ein zentraler Index-Server mehrere, peer-to-peer-basiert verbundene Knoten ansteuert. Sie geben nach einer Implementierung des Ansatzes an, dass die Geschwindigkeit mindestens einem ‚üblichen‘ OLAP-System entspricht und oft höher ist. Sowohl forschungsmethodisches Vorgehen als auch Architekturbeschreibung und -implementierung werden allerdings sehr kurz und wenig transparent dargestellt (vgl. Park/Joo (2013): 43–46).

Die bisher vorgestellten Konzepte basieren auf einer ‚geschickten‘ Verteilung der Daten auf mehrere Systeme, bzw. Datenbanken.<sup>376</sup> AKINDE ET AL. hingegen gehen von einer nicht-veränderlichen Grundstruktur verteilter Data Marts aus und stellen einen Ansatz vor, die OLAP-Abfragen selbst besser zu gestalten und auf die Systeme zu verteilen, um so einen Geschwindigkeitsvorteil, insbesondere vor dem Hintergrund beschränkter Netzverbindungen, zu erzielen. Sie verwenden dazu einen zentralen ‚Coordinator‘, d. h. einen zentralen Server, der die OLAP-Abfragen entgegen nimmt und die Verteilung auf die einzelnen Systeme verwaltet. Sie schaffen damit eine erste Möglichkeit, nicht vollständig integrierte Systeme ohne deren Modifikation unter Berücksichtigung von Optimierungen bei der Abfragegestaltung zu nutzen, schränken dies aber auf eine ‚kleine bis mittlere‘ Anzahl verwalteter Systeme ein und schlagen vor, den koordinierenden Server selbst wiederum auf mehrere Systeme zu verteilen.<sup>377</sup>

Eine noch weniger stark regulierte und vollständig dezentrale Lösung bieten DOKA/TSOUMAKOS/KOZIRIS mit dem ‚Brown Dwarf‘.<sup>378</sup> Sie beziehen sich auf ein P2P-Netzwerk, setzen dafür aber wiederum eine Veränderlichkeit der angeschlossenen Systeme voraus – was im Vergleich zu dem vorherigen Ansatz eine Einschränkung bedeutet. Ziel des Konzepts ist die Verteilung eines originär zentralen DWH auf ein unstrukturiertes ‚Overlay‘<sup>379</sup>, das eine effiziente Abfrageunterstützung auch bei ausfallenden Knoten garantiert.<sup>380</sup> Die Autoren zeigen, dass der Einsatz eines P2P-Netzwerks dabei im Vergleich zu dem auf einer zentralen Struktur basierenden System (‚Dwarf‘) signifikante Geschwindigkeitsvorteile bringen kann, da sowohl Würfel-Erstellung als auch Abfrage-Beantwortung von mehreren Knoten gleichzeitig errechnet werden können. Es liegt damit aus Sicht

---

<sup>376</sup> Ergänzend sei der Beitrag von MAHBOUBI/DARMONT genannt, der die Konzepte relationaler Partitionierung auf den Kontext XML-basierter DWH-Systeme überträgt und den Einsatz der nach eigenen Angaben bisher nicht sehr performanten Systeme mit einer steigenden Heterogenität der für Entscheidungsprozesse notwendigen Daten begründet (vgl. Mahboubi/Darmont (2009): 1555f.).

<sup>377</sup> Vgl. Akinde et al. (2003): 111f., 116f., 132f.

<sup>378</sup> Vgl. Doka/Tsoumakos/Koziris (2011): 1434f.

<sup>379</sup> Der englische Begriff ‚overlay‘ ist hier am ehesten mit ‚Überlagerung‘ oder ‚Überdeckung‘ zu übersetzen. Beide Begriffe geben den Charakter eines die vorhandenen Systeme überspannenden, neu geschaffenen Netzwerks aber nur unzureichend wieder, sodass hier weiterhin der Originalbegriff verwendet wird. Vgl. hierzu auch die Ausführungen zu Schichtenmodellen in Abschnitt 4.5.2.1.

<sup>380</sup> Vgl. Doka/Tsoumakos/Koziris (2011): 1436.



paralleler Verarbeitung hier die unabhängigste Architektur eines verteilten DWH in der vorgestellten Literatur vor.

#### 4.2.2 Peer-to-peer-basiertes Online Analytical Processing

Keine der im vorherigen Abschnitt vorgestellten Lösungen liefert ein vollständiges Modell eines unabhängigen, verteilten, auf unveränderlichen Strukturen basierenden Systems. Dennoch zeigen die bisherigen Ansätze, dass die technischen Möglichkeiten für verteilte Speicherung und Verarbeitung gegeben und erprobt sind. Während ein P2P-basiertes DWH als Gesamtkonzept selbst allerdings nur selten thematisiert wird, existieren mehrere Ansätze, die OLAP-Abfragen über P2P-Netzwerke behandeln. Es lassen sich dabei zwei Verständnisse eines P2P-OLAP-Systems unterscheiden. Im ersten existieren DWH-Knoten, die Daten vorhalten und Abfrage-Knoten, die Daten anfordern und/oder weiterleiten.<sup>381</sup> Im zweiten können alle Knoten sowohl Datenhaltung als auch Abfragegenerierung und Ergebnisweiterleitung übernehmen.<sup>382</sup> Auch wird in beiden Fällen nicht immer ein voll ausgebildetes DWH vorausgesetzt, sondern zumeist nur ein als multidimensional betrachtbarer Datenbestand in einer Datenbank, in der Regel in relationaler Form. In Bezug auf die vorliegende Arbeit sind die im Folgenden vorgestellten Beiträge insbesondere für die Bewahrung der Autonomie der beteiligten Unternehmen relevant,<sup>383</sup> sodass die Erkenntnisse hier unabhängig von dem beschriebenen Netzwerktyp aufgenommen werden. Die Performance eines OLAP-Systems, das mittels eines P2P-Netzwerks angesprochen wird, ist in erheblichem Maße von der Geschwindigkeit des Netzwerks abhängig, da die Übertragung einer hohen Menge an Daten in der Regel mehr Zeit in Anspruch nimmt als die Ausführung der Abfragen.<sup>384</sup> Zur Reduzierung der Laufzeiten über das Netzwerk schlagen KALNIS ET AL. vor, Abfrageergebnisse nicht nur weiterzuleiten, sondern auch temporär für zukünftige, identische Abfragen vorzuhalten („Caching“).<sup>385</sup>

---

<sup>381</sup> Vgl. z. B. Kalnis et al. (2002):27.; Seshadri/Cooper/Liu (2005): 3f.

<sup>382</sup> Vgl. z. B. Vaisman/Minuto Espil/Paradela (2009): 232; Wu et al. (2009): 443.

<sup>383</sup> Vgl. Abschnitt 2.4.

<sup>384</sup> Vgl. Kalnis et al. (2002): 25. Aufgrund der technischen Entwicklung seit 2002 und der starken und vergleichsweise kostengünstigen Verbreitung von Breitbandverbindungen, kann diese Aussage diskutiert werden. Unter der Annahme, dass ein aktuelles, gut konfiguriertes OLAP-System nach den Regeln von CODD/CODD/SALLEY in wenigen Sekunden eine Abfrage liefert, das Ergebnis aber je nach Abfrage mehrere Megabytes umfassen kann, scheint die Annahme zumindest in abgeschwächter Form weiterhin gerechtfertigt. Dies gilt insbesondere, wenn die Ausführung auf den beteiligten Peers (teil-)sequenziell erfolgt.

<sup>385</sup> Vgl. Kalnis et al. (2002): 28f.

Abb. 30 stellt die Grundstruktur des Netzwerks und den vereinfachten Aufbau eines Abfrage-Knotens dar.

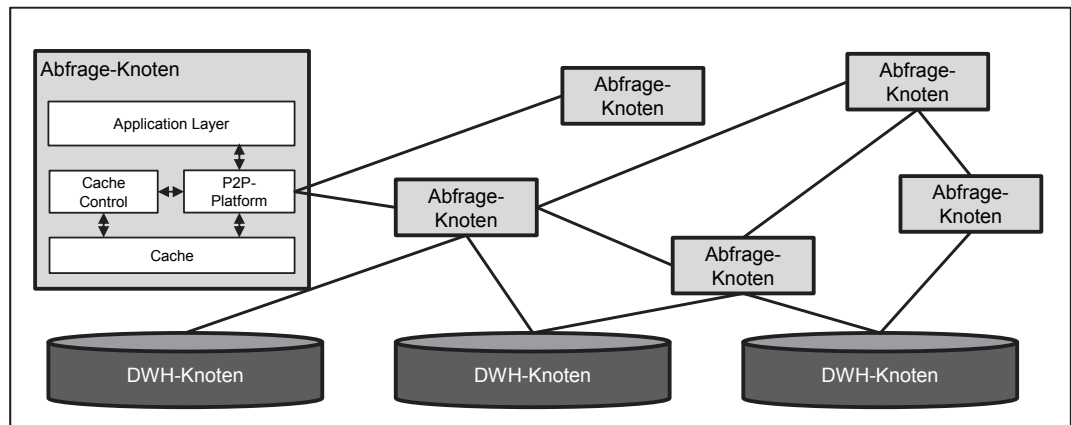


Abb. 30: Schematischer Aufbau eines P2P-OLAP-Netzwerks mit unterschiedlicher Knoten-Definition

(Quelle: In Anlehnung an Kalnis et al. (2002): 28f.<sup>386</sup>)

Die Autoren zeigen, dass sich durch den verteilten Zugriff ein Geschwindigkeitsvorteil gegenüber alternativen Caching-Varianten (ein zentraler Cache, zentrale Caches je DWH-Knoten, kein Caching) erzielen lässt, verweisen jedoch auf die Notwendigkeit einer großen Ähnlichkeit der einzelnen Abfragen, um die Vorteile überhaupt nutzen zu können.<sup>387</sup>

SESHADRI/COOPER/LIU erweitern den vorgestellten Aufbau um zwei Aspekte. Zum einen verbessern sie die Effizienz bei der Abfrage an andere Abfrage-Knoten, indem mit Hilfe von „Query Trails“<sup>388</sup> nachvollzogen wird, welche Nachbarknoten zuletzt ähnliche Abfragen gestellt haben und vermutlich am ehesten die aktuelle Abfrage bedienen können. Zum anderen schaffen sie die Möglichkeit, die Abfrage nach einem aggregierten Wert in Einzelabfragen nach den hierarchisch weiter unten liegenden, ggf. auch aggregierten, Werten aufzusplitten und bereits vorhandene Teilergebnisse aus anderen Knoten zu nutzen. Dies setzt eine unabhängige Aggregationsform der Teilergebnisse, in der Regel die Summation, voraus.<sup>389</sup>

<sup>386</sup> Die Grafik stellt die von KALNIS ET AL. vorgeschlagenen Index-Knoten zur Initialisierung neuer Peers nicht dar, da die Autoren selbst darauf verweisen, dass diese nicht Konzeptnotwendig sind, sondern nur zur praktischen Umsetzung eingebaut wurden (vgl. Kalnis et al. (2002): 28).

<sup>387</sup> Vgl. Kalnis et al. (2002): 35f.

<sup>388</sup> Seshadri/Cooper/Liu (2005): 1.

<sup>389</sup> Vgl. Seshadri/Cooper/Liu (2005): 3–9.

Eine weitere Reduktion der übertragenen Daten in Bezug auf das Gesamtnetzwerk und damit eine Geschwindigkeitsverbesserung erzielen KEHUA/MANIRAKIZA. Sie ergänzen das vorgestellte Konzept um einen zentralen Tracker, d. h. eine Instanz, die die aktuellen Caches der im Netzwerk vorhandenen Knoten speichert und darauf basierend die Abfragen leitet. Es ergibt sich dadurch ein merklicher Geschwindigkeitsvorteil.<sup>390</sup> Im Vergleich zu den ‚simpleren‘ Varianten stellt sich allerdings die Frage nach dem Besitzer des Trackers, sowohl in Bezug auf den notwendigen Aufwand zum Betrieb, als auch in Bezug auf den Zugriff auf Daten bzw. die Informationsfreiheit und damit Autonomie der einzelnen Knoten.

Neben den reinen Routing<sup>391</sup>-Überlegungen muss auch eine Auswahl getroffen werden, welche Inhalte in einem Cache begrenzter Größe überhaupt vorgehalten werden können.<sup>392</sup> Die entsprechenden Überlegungen ähneln dabei in der Fragestellung teilweise Konzepten, die für nicht-verteilte Lösungen diskutiert werden. HALEVY stellt unterschiedliche Ansätze zur Auswahl geeigneter Views<sup>393</sup> für die Auswertung von OLAP-Abfragen vor, die möglichst viele Abfragen ohne Veränderung abdecken sollen.<sup>394</sup> Die gleiche Zielsetzung verfolgen die Auswahlstrategien für ‚gecachte‘, also vorgehaltene Ergebnisse von Abfragen in P2P-Netzwerken. In diesem Zusammenhang findet sich auch die Idee der Bestimmung von Ähnlichkeiten der Abfragen wieder, die bei der Auswahl der anzusprechenden Peers in einem P2P-Netzwerk mit Caching vorgestellt wurde. Ein entsprechender Ansatz findet sich bei AOUICHE/JOUVE/DARMONT, die ein Ähnlichkeitsmaß für einzelne Abfrage-Bestandteile und, davon abgeleitet, für komplette Abfragen entwickeln.<sup>395</sup>

---

<sup>390</sup> Vgl. Kehua/Manirakiza (2012): 700.

<sup>391</sup> *Routing* beschreibt das Weiterleiten von Daten (oder Nachrichten) in einem Netzwerk. Eine Routing-Strategie versucht in der Regel die Kosten (abhängig von der jeweils passenden Definition können dies z. B. die Weglänge, die Belastung der Knoten, etc. sein) der Weiterleitung über die gesamte Strecke zu minimieren (vgl. Aberer et al. (2005): 15).

<sup>392</sup> Vgl. Kehua/Manirakiza (2012): 699f.

<sup>393</sup> Eine *View* ist das Ergebnis einer Datenbankabfrage und stellt damit im Kontext relationer Datebanken selbst eine Relation dar. Wird diese Ansicht nur als Definition vorgehalten, so liegt eine *Virtual View* vor, die bei jeder Abfrage auf die View erneut selbst zunächst erzeugt werden muss. Wird die View hingegen gespeichert, so liegt eine *Materialized View* vor, die den gleichen Zugriffseigenschaften unterliegt wie die eigentlichen Datenquellen, entsprechend aber auch redundante Daten vorhält und Speicherplatz beansprucht (vgl. Cannataro et al. (2006): 729f.).

<sup>394</sup> Vgl. Halevy (2001): 270–275.

<sup>395</sup> Vgl. Aouiche/Jouve/Darmont (2006): 84–87.

Einen verwandten Ansatz liefert SAPIA, der Markov-Modelle<sup>396</sup> verwendet, um zukünftige Abfragen basierend auf bisherigem Nutzungsverhalten abzuleiten. Er bezieht sich dabei auf Operationen auf der aktuellen Abfrage, also z. B. ‚Drill-Down‘ oder ‚Pivoting‘.<sup>397</sup> Eine Übertragung des Vorgehens auf P2P-Netzwerke könnte z. B. die Auswahl des nächsten Peers oder die Caching-Strategie bei wenig Speicherplatz beeinflussen.<sup>398</sup> Eine weiterführende Übersicht zur Ähnlichkeitsbestimmung von OLAP-Abfragen und vollständigen OLAP-‚Sessions‘, d. h. geordneten Abfragereihenfolgen, findet sich bei ALIGON ET AL.<sup>399</sup>

Als Alternative zu Caching-Strategien für die exakte Abfrage von Ergebnissen präsentieren WU ET AL. einen Ansatz, der auf Näherungen von Ergebnissen basiert. Eine gestellte Abfrage wird an einen Datenknoten geleitet, welcher kontinuierlich wachsende Datenauszüge an die Peers des Abfrage-Knotens schickt. Diese berechnen parallel Näherungen für Teilergebnisse, die dann im originären Abfrage-Knoten wieder kombiniert werden und mit Angabe eines Konfidenz-Intervalls angezeigt werden. Der Prozess wird solange ausgeführt, bis entweder das exakte Ergebnis vorliegt oder ein voreingestelltes Konfidenz-Intervall erreicht wird.<sup>400</sup>

Ein umfangreiches P2P-OLAP-Konzept liefern VAISMAN/MINUTO E-SPIL/PARADELA. Im Gegensatz zu den bisher vorgestellten Ansätzen gehen sie nicht davon aus, dass alle Peers mit dem gleichen Schema arbeiten, sondern betrachten die einzelnen Peers als autonome Datenquellen, die jedoch immer nur mit

---

<sup>396</sup> Ein *Hidden Markov Model* basiert auf einer Menge von Zuständen, die aufeinander überführbar sind, wobei – hier sei aus Gründen der Einfachheit auf eine vollständige formale Definition verzichtet – auf einen Zustand  $s_i$  mit einer Wahrscheinlichkeit  $p_{ij}$  der Zustand  $s_j$  folgt, die Übergangswahrscheinlichkeit konstant sowie von vorherigen Zustandsänderungen unabhängig ist und die einzelnen Zustände nicht beobachtbar sind. Durch mit einer gegebenen Wahrscheinlichkeit erzeugte Emissionen lässt sich aber ein Rückschluss auf den wahrscheinlichen Zustand des Modells und seiner Entwicklung treffen. Für weiterführende Erklärungen vgl. z. B. Rabiner (1988).

<sup>397</sup> Vgl. Sapia (2000): 227–229.

<sup>398</sup> Es ist zu vermuten und stimmt mit den bisherigen Erkenntnissen aus der Literatur überein, dass auf Grund des Alters der Publikation ein P2P-OLAP-System zu Beginn der Modellentwicklung des Autors nicht so populär war, dass eine Übertragung nahelegen hat.

<sup>399</sup> Vgl. Aligon et al. (2014b): 466–471. Der Ansatz wird in einem fortführenden Beitrag erweitert und zur Vorschlagsgenerierung für OLAP-Anwender verwendet (vgl. Aligon et al. (2014a): 2f.). Die Autoren beschäftigen sich in ihrem Beitrag auch mit der Methodik der Entwicklung von Ähnlichkeitsmaßen. Auch wegen der thematischen Nähe wird der Beitrag in Abschnitt 5.2.2 dieser Arbeit wieder aufgegriffen.

<sup>400</sup> Vgl. Wu et al. (2009): 443–445.

direkt verbundenen Knoten kommunizieren können.<sup>401</sup> Es entsteht so kein Routing-Problem, wohl aber, wie bei den anderen Konzepten, ein Performancebestimmendes Caching-Problem.<sup>402</sup> Im Fokus stehen bei den Autoren die durch die Schemavielfalt entstehenden Fragestellungen. Sie betreffen die Überführbarkeit der Abfragen und Interpretierbarkeit der Abfrageergebnisse. Da die Annahme (auch fachlich) autonomer Netzwerkteilnehmer zentral für die vorliegende Arbeit ist, setzt sich der folgende Abschnitt ausführlich mit der Thematik heterogener Schemata auseinander.

### 4.3 Harmonisierung und Interpretierbarkeit multidimensionaler Datenstrukturen

Für die Zusammenführung unterschiedlicher multidimensionaler Datenbestände mit originär unterschiedlichen Schemata lassen sich drei grundlegende Ansätze identifizieren:<sup>403</sup>

- In einem *zentralen DWH mit globalem/einheitlichem Schema* werden die Daten der Quellen zusammengeführt und in der Regel auch passend materialisiert. Dies entspricht im Wesentlichen den Überlegungen zum Aufbau von (zentralen) DWH-Systemen, wie sie in den Abschnitten 3.1.2 und 4.2.1 vorgestellt wurden.
- Ein *verteiltes DWH mit globalem/einheitlichem Schema* („federated DWH“) belässt die einzelnen Datenquellen physisch autonom und muss Ansätze finden, wie Abfragen über mehrere Datenquellen verteilt werden können. Im Sinne getrennter Data Marts als Bestandteile eines DWH entspricht dies dem föderierten Ansatz aus Abschnitt 4.2.1.
- *Unabhängige DWH-Systeme mit jeweils eigenen Schemata* bilden die loseste Kopplung verschiedener Systeme. Abfragen an den Gesamtdatenbe-

---

<sup>401</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 231f. Eine frühere Version der Arbeit beschreibt bereits einige Jahre zuvor das entsprechende Konzept, teilweise sogar mit den gleichen Formulierungen und Grafiken (vgl. Minuto Espil/Vaisman (2004)). GOLFARELLI ET AL. verweisen auf die erstgenannte Arbeit und stellen als essenziellen Unterschied zu der ihren ebenfalls die Notwendigkeit eines vollständig verbundenen Netzes heraus, was dazu führt, dass Fragestellungen, die typisch für Netze sind, die nicht dieser Eigenschaft genügen, nicht diskutiert werden. Dazu gehört insbesondere der Umgang mit einem nicht global vorhandenen Wissen über alle Schemata und der daher notwendigen mehrfachen Abfrage-Umformulierung (vgl. Golfarelli et al. (2012b): 395f.).

<sup>402</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 257.

<sup>403</sup> Vgl. hierzu und zu der folgenden Aufzählung Rizzi (2012): 190–193; Tseng/Chen (2005): 210.

stand müssen zunächst auf die einzelnen Schemata verteilt werden oder direkt auf die entsprechenden Schemata ausgerichtet sein.

Diese Einteilung sowie die zugrunde liegenden Überlegungen sind durch die bisherigen Aufarbeitungen für den Ein-Unternehmen-Fall in weiten Teilen erklärt. Eine Betrachtung des Mehr-Unternehmen-Falls, wie ihn die CCBI vorsieht, ändert zunächst die Fragestellung der Herstellung eines integriert auswertbaren Datenbestandes nicht, begünstigt aber die Idee unabhängiger Systeme, da die geforderte Berücksichtigung der Autonomie der Beteiligten hier am offensichtlichsten erzielt wird. Die folgenden Abschnitte stellen daher zunächst die allgemeinen Probleme und grundlegenden Lösungsansätze bei der Integration unterschiedlicher Datenbanken vor. Im Folgenden wird auf die Spezifika im DWH-Kontext eingegangen. Dabei wird auch die Vielzahl an Integrationsansätzen im Bereich der Ontologieforschung berücksichtigt, die Ähnlichkeiten insbesondere zu den Dimensionen im DWH-Kontext aufweisen.<sup>404</sup> Abschließend folgt eine Betrachtung der Ansätze, die besonders auf unabhängige Datenbasen und Netzwerke eingehen.

#### 4.3.1 Konflikte unterschiedlicher multidimensionaler Datenstrukturen

Nach BERGER sind multidimensionale Datenstrukturen grundsätzlich semantisch ‚reicher‘ als typischerweise in Datenbanken abgelegte relationale Datenbestände, da immer auch eine Unterscheidung zwischen Dimensionen und Fakten vorliegt.<sup>405</sup> Dies führt dazu, dass bereits bei dem Aufbau eines einzelnen Würfels ein Konfliktpotenzial zwischen unterschiedlichen Modellierungsvarianten existiert. Dieses Potenzial steigt aber weiter an, wenn unterschiedliche Würfel, die sich mit dem gleichen Sachverhalt befassen, miteinander verglichen werden bzw. zusammengeführt werden sollen, da in diesem Fall unterschiedliche Design-Entscheidungen für die Schemata – und auch die Instanzen<sup>406</sup> – getroffen werden können.

Ein Beispiel einer problematischen Design-Entscheidung – bereits bei einem Würfel – ist die Verwendung heterogener Dimensionen. Diese kennzeichnen sich dadurch, dass nicht alle Elemente einer Ebene zu Elementen der gleichen höheren

---

<sup>404</sup> Vgl. Abschnitt 3.3.3.1.

<sup>405</sup> Vgl. Berger (2009): 60.

<sup>406</sup> Zur einfachen sprachlichen Verwendung bezeichnet das *Schema* einer Dimension die verwendeten Ebenen mit ihren jeweiligen Bedeutungen, eine *Instanz* die tatsächlichen Ausprägungen der Elemente in den Ebenen. Für die formale Definition vgl. Abschnitt 3.1.3. Das Verständnis ist entsprechend auf Würfel zu übertragen.

Ebene aggregiert werden.<sup>407</sup> Abb. 31 zeigt die Instanz einer heterogenen Dimension basierend auf dem bereits bekannten Beispiel. Die Summe aller Werte, die der zweiten Ebene zugerechnet werden können, ist jetzt nicht mehr identisch mit dem aggregierten Wert des Wurzelknotens der Hierarchie („Welt“). Eine Aufspaltung von Abfragen an einen OLAP-Würfel mit einer solchen Dimension kann dadurch nicht mehr gewährleisten, dass das Gesamtergebnis korrekt wiedergegeben wird, was Strategien für verteiltes Data Warehousing oder P2P-OLAP stark beeinträchtigt.<sup>408</sup>

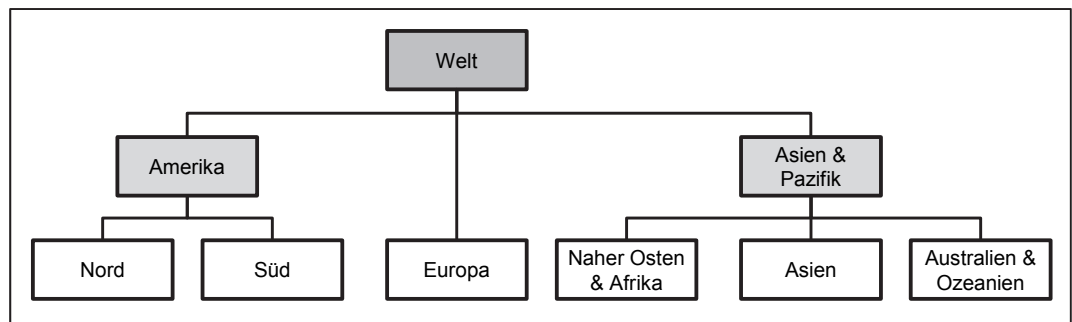


Abb. 31: Beispielhafte Instanz einer heterogenen Hierarchie

(Quelle: Eigene Darstellung)

Bezogen auf die Kombination zweier multidimensionaler Datenstrukturen, d. h. Würfel, Data Marts, etc., lassen sich fünf Konfliktarten unterscheiden, die abhängig davon sind, ob Instanzen oder Schemata verglichen werden und nur einzelne Dimensionen oder vollständige Würfel, insbesondere mit Verweis auf die Faktentabelle, betrachtet werden (vgl. Abb. 32).

<sup>407</sup> Vgl. Hurtado/Mendelzon (2002): 169.

<sup>408</sup> Zu ausführlichen Beispielen und weiteren Varianten der vorgestellten Komplikation vgl. Mazón/Lechtenbörger/Trujillo (2009): 1453–1462.

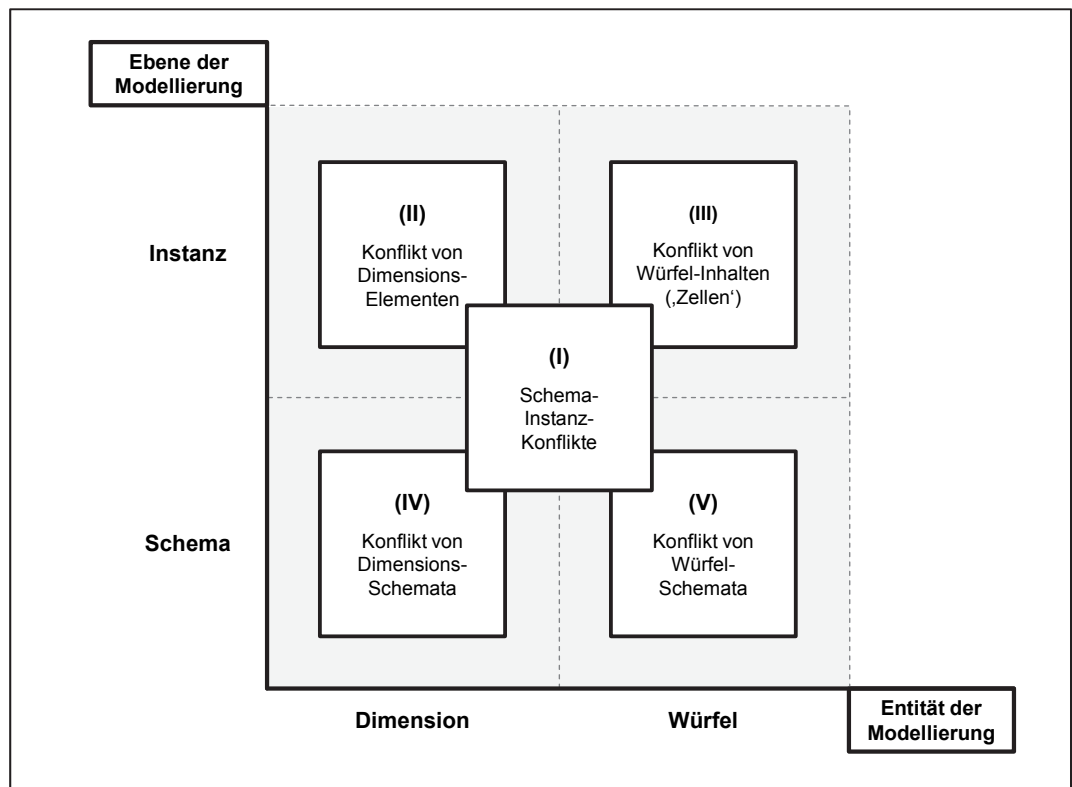


Abb. 32: Klassifikation möglicher Konflikte bei multidimensionalen Datenstrukturen

(Quelle: In Anlehnung an Berger (2009): 61)

*Schema-Instanz-Konflikte* (in der Abbildung mit I gekennzeichnet) entstehen, wenn der gleiche Sachverhalt abgebildet wird, dazu aber unterschiedliche Konzepte verwendet werden, die die Semantik der gespeicherten Werte einmal als Teil des Schemas, einmal als Teil der Instanz vorhalten.<sup>409</sup> Ein in der Praxis typisches Beispiel ist die gegensätzliche Modellierung z. B. von Geschäftszahlen im ‚Kennzahlenmodell‘ oder ‚Kontenmodell‘. Im ersten Fall enthält die Faktentabelle zu jeder Dimensionselementkombination mehrere Fakten, die semantisch beschrieben werden (z. B. als ‚Umsatz‘ und ‚Menge‘); im zweiten Fall wird generell nur ein Fakt pro Elementkombination vorgehalten, und die semantische Bedeutung über eine zusätzliche Dimension abgebildet (vgl. die beispielhafte Darstellung in Abb. 33).<sup>410</sup>

<sup>409</sup> Vgl. Berger (2009): 62.

<sup>410</sup> Gemäß der in Abschnitt 3.1.3 eingeführten Notation wird das Faktenschema  $f[A_1:l_1, \dots, A_n:l_n] \rightarrow \langle M_1:\tau_1, \dots, M_m:\tau_m \rangle$  bei einem Kennzahlenmodell wie vorgestellt verwendet – unter der Annahme, dass es  $n$  tatsächlich beschreibende Dimensionen und  $m$  Kennzahlen gibt –, während für ein korrespondierendes Kontenmodell eine weitere Dimension ( $n + 1$ ) eingeführt würde, die die Beschreibung der Kennzahlen enthält und  $m = 1$  gilt.



Kennzahlenschema				Kontenschema			
Dimension 1 'Produkt'	Dimension 2 'Geographie'	Kennzahl 1 'Umsatz'	Kennzahl 2 'Menge'	Dimension 1 'Produkt'	Dimension 2 'Geographie'	Dimension 3 'Kennzahl'	Kennzahl- wert
Produkt 1	Europa	1000	50	Produkt 1	Europa	Umsatz	1000
Produkt 2	Asien	2000	30	Produkt 1	Europa	Menge	50
...	...	...	...	Produkt 2	Asien	Umsatz	2000
				Produkt 2	Asien	Menge	30
				...	...	...	...

Abb. 33: Kennzahlen- und Kontenschema im Vergleich an einem Beispiel

(Quelle: In Anlehnung an Mehrwald (2010): 53)

*Schema-Level-Konflikte* (IV, V) entstehen z. B. bei unterschiedlichen Namenskonventionen, sodass Synonyme oder Homonyme entstehen, was zunächst kein DWH-spezifisches Problem darstellt. Ähnliche Probleme, z. B. durch unterschiedliche Elemente mit identischen Beschreibungen, entstehen auch bei *Instanz-Level-Konflikten* (II, III). Die folgende Auflistung stellt zur Verdeutlichung prägnante Beispiele der jeweiligen Konfliktklasse dar, wobei eine Gegenüberstellung zweier Würfel  $W_1$  und  $W_2$  angenommen wird.<sup>411</sup>

- (II) Ein Element ist in einer Dimension sowohl in  $W_1$  als auch in  $W_2$  vorhanden, jedoch unterschiedlich in das Hierarchiegefüge eingebunden, d. h. die Elemente liegen nicht auf einer korrespondierenden Ebene.
- (III) Ein Fakt wird durch abweichende Attributkombinationen von  $W_1$  und  $W_2$  beschrieben oder identische Attributkombinationen verweisen auf unterschiedliche Kennzahlen.
- (IV) Dimensionen sind unterschiedlich definiert. So ist z. B. eine Dimension ‚Zeit‘ in  $W_1$  durch die Hierarchie ‚Tag  $\leq$  Woche  $\leq$  Jahr‘ beschrieben, in  $W_2$  durch die Hierarchie ‚Tag  $\leq$  Monat  $\leq$  Jahr‘.
- (V) Die Fakten stellen zwar grundsätzlich einen identischen Sachverhalt dar, weisen aber unterschiedliche Semantik auf, z. B. sind Umsatzwerte zum einen als Euro-, zum anderen als US-\$-Werte zu interpretieren.

Für alle vorgestellten Konfliktpotenziale können unterschiedlich schwere Probleme auftreten, die jedoch zunächst überhaupt eine Zusammenführung der Datenbanken erfordern. Vor dem Hintergrund einer relationalen Abbildbarkeit multidimensionaler Datenstrukturen werden im Folgenden daher zunächst – zumeist relationalen Datenbanken entlehnte – Ansätze zur Schemaintegration vorgestellt.

<sup>411</sup> Vgl. Berger (2009): 63–75.

### 4.3.2 Grundlegende Lösungsansätze für verteilte Datenbasen

Die Integration relationaler Datenbankschemata ist ein umfangreich erforschtes Gebiet. Gesucht wird für zwei Schemata ein *Match*, d. h. eine Zuordnung („mapping“) der Elemente zweier Schemata, sodass eine semantisch korrekte Überführung stattfindet.<sup>412</sup> RAHM/BERNSTEIN liefern eine vielzitierte Übersicht zu bekannten Matchingverfahren und bilden eine Klassifikation dieser. Sie unterscheiden die Ansätze nach dem Hauptkriterium „Instance vs schema“<sup>413</sup>. Dabei wird betrachtet, ob ein Schema-Matchingverfahren – aber nicht ein Instanz-Matchingverfahren – auch die Elemente beispielhafter Instanzen berücksichtigt oder lediglich die Informationen zum Schema selbst zu Grunde gelegt werden. Weitere Kriterien beziehen sich auf die Komplexität der Strukturen im Vergleich, den Einsatz linguistischer Werkzeuge, die Berücksichtigung von Kardinalitäten der Beziehungen von Elementen und den Einsatz weiterer Hilfsmittel, denen auch auch Benutzereingaben zuzurechnen sind.<sup>414</sup>

Eine neuere Übersicht von Verfahren mit einer erweiterten Klassifikation stammt von SHVAIKO/EUZENAT. Die Autoren unterscheiden dabei nicht nur nach dem Betrachtungsobjekt (Schema und/oder Instanz), sondern auch nach der Art der verwendeten Methoden zur Abdeckung z. B. linguistischer Übereinstimmungen. Insbesondere wird der Einbezug semantischer Abgleiche berücksichtigt, wie er durch den Einsatz von Wortdatenbanken – u. a. für das Finden von Synonymen – ermöglicht wird.<sup>415</sup> Die Klassifikation unterscheidet auch zwischen reinem Schema-Matching und Ontologie-Matching und zieht als entscheidendes Kriterium den Grad der Semantik heran, die bei einer Ontologie erheblich stärker ausgeprägt ist als bei einer reinen Schemabeschreibung.<sup>416</sup>

---

<sup>412</sup> Vgl. Rahm/Bernstein (2001): 334; sowie die formale Aufarbeitung in Cate/Kolaitis (2010). Die Unterscheidung von ‚mapping‘ und ‚matching‘ wird in der Literatur nicht einheitlich vorgenommen (vgl. Bellahsene et al. (2011): 254). In der vorliegenden Arbeit wird – soweit es den Originalquellen nicht widerspricht – ein ‚Mapping‘ verstanden als die Zuordnung oder Überführungsregel(n) zweier Schemata, ein ‚Matching‘ hingegen als der Prozess oder das Verfahren zur Schaffung dieses Mappings.

<sup>413</sup> Rahm/Bernstein (2001): 337.

<sup>414</sup> Vgl. Rahm/Bernstein (2001): 337f.

<sup>415</sup> Vgl. Shvaiko/Euzenat (2005): 154f. Eine oft genannte Quelle ist WordNet (<http://wordnet.princeton.edu/>), die eine umfangreiche, aufbereitete Datenbank der englischen Sprache beinhaltet (vgl. Fellbaum (2005): 665–670).

<sup>416</sup> Vgl. Shvaiko/Euzenat (2005): 148f. Zu den Ontologie-basierten Verfahren vgl. Abschnitt 4.3.3.

Unabhängig von der Durchführung des Matchings im Sinne der Mapping-Kriterien, werden zwei Ansätze zur Abbildung der Integration mehrerer einzelner Schemata in ein übergeordnetes Schema betrachtet, die in der Literatur als *global-as-view* (GAV) und *local-as-view* (LAV) bezeichnet werden.<sup>417</sup> Dabei wird in beiden Fällen zunächst ein globales Schema entwickelt, welches bei Verwendung des GAV-Ansatzes üblicherweise materialisiert wird, bei Verwendung des LAV-Ansatzes jedoch nur als gemeinsamer Nenner unabhängiger Quellen fungiert. Bei der Verwendung des GAV-Ansatzes wird eine View<sup>418</sup> relationaler lokaler Datenbanken gebildet, die die lokalen Quellen verknüpft. Daten, die nicht in den lokalen Datenbankenschemata enthalten sind, werden im globalen Schema entsprechend mit *NULL*-Werten<sup>419</sup> befüllt. Eine Abfrage auf das globale Schema kann mit einfachen Mitteln auf die lokalen Schemata heruntergebrochen werden, berücksichtigt aber immer nur die global definierten Eigenschaften. Es reagiert nicht auf Veränderungen in lokalen Schemata. Im Gegensatz dazu wird beim LAV-Ansatz jede Quelle als View auf ein globales Schema definiert. Dabei werden nur die lokalen Datenquellen materialisiert vorgehalten. Abfragen an das globale Schema können lokale Gegebenheiten mit einschließen – soweit diese bekannt sind. Die Abfragen sind erheblich komplexer und anfälliger hinsichtlich einer performanten Verarbeitung, weil zunächst herausgefunden werden muss, welche Kombination lokaler Quellen die Abfrage überhaupt abdeckt.<sup>420</sup> Zur Überwindung der Einschränkungen beider Ansätze wurde als Spezialfall des LAV-Ansatzes *global-local-as-view* (GLAV) entwickelt, bei dem korrespondierende Sichten auf die lokalen Schemata gebildet, aber nicht materialisiert werden.<sup>421</sup>

Im Vergleich dazu umfassender ist der Ansatz *both-as-view* (BAV), bei dem komplette *Transformationspfade* zwischen einem globalen Schema und den lokalen Schemata definiert werden, sodass unabhängig von der Materialisierung eine Änderung sowohl des globalen als auch der lokalen Schemata möglich ist.<sup>422</sup> Die-

---

<sup>417</sup> Vgl. Golfarelli et al. (2012b): 396.

<sup>418</sup> Vgl. Abschnitt 4.2.2.

<sup>419</sup> Ein *NULL*-Wert entspricht der Information, dass an der bezeichneten Stelle kein Datum vorliegt, entweder, weil die beschreibenden Attribute der Stelle nicht zulässig sind oder der Wert (derzeit) nicht bekannt ist (vgl. Codd (1979): 403).

<sup>420</sup> Vgl. McBrien/Poulovassilis (2003): 227–229.

<sup>421</sup> Vgl. Cali et al. (2003): 349.

<sup>422</sup> Vgl. McBrien/Poulovassilis (2003): 3–5, 9–11

ser Ansatz wurde für den Einsatz in P2P-Umgebungen verfeinert und wird bei der Betrachtung netzbasierter Ansätze wieder aufgegriffen.<sup>423</sup>

### 4.3.3 Data-Warehouse-spezifische und Ontologie-basierte Ansätze

Die Überführung multidimensionaler Datenstrukturen weist im Vergleich zu den ‚flachen‘ Strukturen einfacher Datenbanken sowohl erschwerende als auch vereinfachende Spezifika auf.<sup>424</sup> Die semantische Trennung in Dimensionen und Fakten sowie die Berücksichtigung der korrekten Aggregationen – z. B. bei Instanz-Konflikten auf Dimensionsebene – kann nicht einfach durch die bisherigen Ansätze (GAV, LAV, GLAV) abgedeckt werden.<sup>425</sup> Erleichternd liegt bei den diskutierten multidimensionalen Objekten allerdings die Dimensions-Fakt-Struktur erwartbar vor und die Daten sind auf Grund vorheriger Aufarbeitungsprozesse bei der Aufnahme in das originäre DWH bereits in einheitlichem Format, was eine systematische Betrachtung der Matching-Voraussetzungen und -Prozesse erlaubt.<sup>426</sup> TORLONE erarbeitet drei erstrebenswerte Eigenschaften eines Dimensions-Matchings, die, wenn sie alle erreicht werden, ein ‚perfektes‘ Mapping definieren.<sup>427</sup>

**Definition (7)** Ein *Dimensions-Matching* ist eine Funktion  $\mu: d_1 \rightarrow d_2$ , die den Hierarchieebenen von  $d_1$  Hierarchieebenen von  $d_2$  zuordnet.

**Definition (8)** *Coherence* liegt vor, wenn für jede Paarung von Hierarchieebenen  $l, l'$  die Aggregationsordnung auch nach Überführung erhalten bleibt, d. h.  $l \preccurlyeq_1 l' \leftrightarrow \mu(l) \preccurlyeq_2 \mu(l')$  gilt.

---

<sup>423</sup> Vgl. Abschnitt 4.3.4.

<sup>424</sup> Vgl. Bergamaschi et al. (2011): 38.

<sup>425</sup> Vgl. Golfarelli et al. (2012b): 396.

<sup>426</sup> Vgl. Torlone (2008): 70.

<sup>427</sup> Vgl. zur folgenden Aufzählung Torlone (2008): 77f. Ein Dimensions-Matching überführt ein Dimensions-Schema in ein anderes. Wie bisher werden die englischen Begriffe beibehalten und in deutsche Notation überführt. Die formale Notation folgt der in Abschnitt 3.1.3 eingeführten Dimensionsalgebra.

**Definition (9)** *Soundness* liegt vor, wenn für jede Überführung ein vollständiges Mapping aller Elemente erreicht wird, d. h.  $m_1(l) = m_2(\mu(l))$  gilt.<sup>428</sup>

**Definition (10)** *Consistency* liegt vor, wenn die Aggregationsfunktion  $\rho$  für jede Paarung von Hierarchieebenen nach Überführung erhalten bleibt, d. h.  $\rho_1^{l \rightarrow l'} = \rho_2^{\mu(l) \rightarrow \mu(l')}$  gilt.<sup>429</sup>

Der Autor stellt im Folgenden zwei diese Eigenschaften erfüllende Ansätze zur Dimensionsintegration vor. Der *loosely-coupled*-Ansatz sucht nach sich eindeutig überschneidenden Datenmengen und definiert Überführungsregeln ausschließlich auf diesen Mengen, sodass ein ‚kleinster gemeinsamer Nenner‘ erzielt wird.<sup>430</sup> Dieser stellt jedoch üblicherweise keinen Zugriff auf den Gesamtbestand dar und gewährleistet damit auch nicht (sachlich) korrekte Aggregationen höherer Ebenen. Der *tightly-coupled*-Ansatz basiert auf der Materialisierung einer umfassenden Sicht auf die Daten, die ggf. neue oder veränderte Dimensionen beinhaltet, sodass alle Quelldaten als perfektes Matching auf die umfassende Sicht überführbar sind,<sup>431</sup> was die bereits diskutierten Probleme hinsichtlich Autonomie und globaler Schemata mit sich führt.

TORLONE gibt abschließend an, dass die definierten Eigenschaften eine Grundlage für Ontologie-basierte Matching-Verfahren darstellen könnten.<sup>432</sup> Auch andere Autoren weisen auf die Nützlichkeit von Ontologien im DWH-Kontext hin. So beschreiben PARDILLO/MAZON den Einsatz von Ontologien u. a. zur Beschreibung der Meta-Daten im DWH,<sup>433</sup> wie es auch HARTMANN zur Überwindung semanti-

<sup>428</sup> Diese Bedingung bezieht die Instanz-Definition mit ein, was die Eigenschaften implizit nur auf Schema-Matching-Verfahren anwendbar macht, die Instanzen berücksichtigen, auch, wenn dies von den Autoren nicht expliziert wird. Da im DWH-Kontext aber – wie beschrieben – insbesondere die Instanz-Konflikte berücksichtigt werden müssen, können die Eigenschaften für alle DWH-spezifischen Verfahren als geeignet betrachtet werden.

<sup>429</sup> Während Coherence ‚nur‘ die Reihenfolge der Hierarchieebenen einhält, fordert Consistency auch die gleichbleibende Zuordnung der Elemente je Ebene, d. h. die Zuordnungen der Aggregationen müssen nach Überführung einen identischen aggregierten Wert erhalten, sofern die Soundness-Bedingung erfüllt wurde.

<sup>430</sup> Vgl. Torlone (2008): 82–85.

<sup>431</sup> Vgl. Torlone (2008): 85–87.

<sup>432</sup> Vgl. Torlone (2008): 96.

<sup>433</sup> Vgl. Pardillo/Mazon (2011): 84f.

scher Heterogenität vorschlägt.<sup>434</sup> KEHLENBECK/BREITNER verwenden Ontologien zur Beschreibung der Kennzahlen in OLAP-Würfeln.<sup>435</sup> Es wird daher im Folgenden geprüft, welche Konzepte bereits im Bereich der allgemeinen Ontologieforschung existieren und ob sie auf DWH-spezifische Probleme anwendbar sind.

Eine aktuelle Übersicht von Ontologie-Matching-Verfahren bieten SHVAIKO/EUZENAT und definieren Ontologie-Matching zunächst als die Überwindung semantischer Heterogenität durch die Zuordnung korrespondierender Elemente in verschiedenen Ontologien.<sup>436</sup> Im Fokus der Forschung stehen bei den in der Übersicht vorgestellten Algorithmen Ähnlichkeitsbestimmungen von Elementen zur automatischen Schemaüberführung – so wie auch in einer vorherigen Übersicht von ABELS/HAAK/HAHN vorgestellt.<sup>437</sup> Ein früher Ansatz, der mehrere Vergleichsansätze kombiniert, basiert auf dem *Jaccard-Koeffizienten*<sup>438</sup>, der hier die Verteilung der Wahrscheinlichkeiten von Instanzzugehörigkeiten zu zwei verschiedenen Ontologieschemata abbildet.<sup>439</sup> Eine Evaluation verschiedener Ähnlichkeitsmaße nehmen BERNSTEIN ET AL. vor. Sie vergleichen die Ergebnisse automatischer Matching-Verfahren mit einer Kontrollgruppe von 50 von Menschen durchgeführten Mappings und schlussfolgern, dass sowohl Matching-Verfahren als auch menschliche Zuordnungen uneinheitlich sind, die Güte der verwendeten Ähnlichkeitsmaße jedoch teilweise erheblich von der verwendeten Ontologie abhängig ist.<sup>440</sup> Diese Einschätzung wird insofern bestätigt, als eine Kombination verschiedener Ähnlichkeitsmaße fast durchgängig bessere Ergebnisse erzielen konnte.<sup>441</sup> Keines der älteren oder aktuellen Verfahren berücksichtigt allerdings die oben definierten Eigenschaften bei der Zusammenführung von Dimensionen, sodass eine direkte Übertragung von Ontologie-Matching-Verfahren auf die Zusammenführung multidimensionaler Datenstrukturen nicht möglich ist.

---

<sup>434</sup> Vgl. Hartmann (2008): 231.

<sup>435</sup> Vgl. Kehlenbeck/Breitner (2009): 299.

<sup>436</sup> Vgl. Shvaiko/Euzenat (2013): 158.

<sup>437</sup> Vgl. Shvaiko/Euzenat (2013): 160–163; Abels/Haak/Hahn (2005): 76f.

<sup>438</sup> Der Jaccard-Koeffizient betrachtet die Wahrscheinlichkeit, dass zwei Objekte gleiche Ausprägungen besitzen unabhängig von Eigenschaften, die beide Objekte *nie* erfüllen und kann mengentheoretisch beschrieben werden als  $\frac{P(A \cap B)}{P(A \cup B)}$ , mit A, B als zu vergleichenden Mengen (vgl. Eckey/Kosfeld/Rengers (2002): 219; Doan et al. (2004): 391).

<sup>439</sup> Vgl. Doan et al. (2004): 391, 393f. Der Ansatz basiert auf dem für XML-Daten entwickelten Ansatz in Doan/Domingos/Halevy (2003).

<sup>440</sup> Vgl. Bernstein et al. (2005): 1359.

<sup>441</sup> Vgl. Ji/Haase/Qi (2011): 293.

Der Grundgedanke einer Ähnlichkeits-basierten, automatisierten Zusammenführung wird DWH-spezifisch von BANEK ET AL. aufgebracht, die mit Hilfe des ‚similarity-flooding‘-Algorithmus<sup>442</sup> Dimensionstabellen in gerichteten Graphen abbilden und eine Strukturähnlichkeit rekursiv vom Wurzelknoten (d. h. der höchsten Aggregationsstufe) aus bestimmen. Dabei adaptieren sie aus dem Ontologie-Matching den Gedanken semantischer Zusammengehörigkeit und berücksichtigen bei der Berechnung von Knotenähnlichkeiten neben der linguistischen Analyse der zugeordneten Elemente auch die Ähnlichkeiten der Nachbarschaftsknoten.<sup>443</sup> Die Autoren erlauben dabei explizit heterogene Dimensionsstrukturen.

BERGAMASCHI ET AL. nutzen die heterogenen Dimensionsstrukturen in einem weitergehenden Ansatz explizit für die Bestimmung von Ähnlichkeiten. Sie argumentieren, dass bei der Zusammenführung von Datenquellen mit grundsätzlich semantischer Ähnlichkeit, wie es z. B. bei mehreren Data Marts eines Unternehmens der Fall ist, einzelne Dimensionen sehr ähnliche Strukturen haben müssen, weil sie gemeinsame Realwelt-Gegebenheiten abbilden, im Beispiel die Unternehmensstruktur. Die so gefundenen Matchings werden mit Hilfe einer linguistischen Analyse – analog zu BANEK ET AL. – validiert. Die Kombination verschiedener Ansätze zeigt sich auch hier zuverlässiger als ein einzelnes Ähnlichkeitsmaß.<sup>444</sup>

Ebenfalls auf dem ‚similarity-flooding‘-Algorithmus basiert der Ansatz von RIAZATI/THOM/ZHANG, der insbesondere die Einhaltung einer weiteren erstrebenswerten Eigenschaft garantiert:<sup>445</sup>

**Definition (11)** *Strictness* beschreibt innerhalb einer Dimension die Aggregation zu maximal einem Element auf der nächsthöheren zugeordneten Ebene im Aggregationspfad, d. h. für jede Paarung von Hierarchieebenen  $l, l'$  ist  $\rho^{l \rightarrow l'}$  rechtseindeutig, bzw. funktional.<sup>446</sup>

<sup>442</sup> Für eine ausführliche Beschreibung des zu Grunde liegenden Algorithmus vgl. Melnik/Garcia-Molina/Rahm (2002): 4–8.

<sup>443</sup> Vgl. Banek et al. (2007): 50f.

<sup>444</sup> Vgl. Bergamaschi et al. (2011): 39f., 46.

<sup>445</sup> Vgl. hierzu und zur folgenden Definition Riazati/Thom/Zhang (2011): 9f.

<sup>446</sup> Tatsächlich beschreibt TORLONE in der Dimensionsalgebra  $\rho$  als ‚Funktion‘, was eine Rechtseindeutigkeit bereits impliziert. Diese Eigenschaft wird jedoch von ihm im Weiteren nicht thematisiert, sondern aus der Betrachtung ausgeschlossen (vgl. Torlone (2008): 74, 96). Die Einführung einer Explikation dieser Eigenschaft ist daher angemessen, um Algorithmen zu beschreiben, die Strictness berücksichtigen.



Damit wird in den zusammengeführten Dimensionen die Summationsproblematik der Doppelzählung gelöst, nicht jedoch die der Nicht-Zählung in heterogenen Dimensionen.<sup>447</sup> Die Autoren nehmen dabei in Kauf, dass nicht alle Elemente aufeinander überführt werden können, sodass als Ergebnis ein unvollständiges Mapping entsteht. Mehrfache Hierarchien können grundsätzlich nicht berücksichtigt werden, ein menschlicher Eingriff am Ende des Prozesses ist nahezu unvermeidbar.<sup>448</sup>

Eine erste integrierte Umsetzung der vorgestellten DWH-spezifischen Ansätze liefert OLARU. Sein Ansatz basiert auf einer Integration der Dimensionen durch Import von Abweichungen der einen Dimension in die andere, sodass eine neue Dimension geschaffen wird. Der Autor verweist darauf, dass die automatische Generierung dieser neuen Dimensionen entscheidend von der Güte eines vorherigen Mappings abhängt.<sup>449</sup> Als Alternative zu einem integrierten Vorgehen kann die Beibehaltung sämtlicher Hierarchien in den Dimensionen verstanden werden, sodass eine ‚Haupthierarchie‘ als kleinster gemeinsamer Nenner existiert und Nebenhierarchien für einzelne Datenquellen gebildet werden.<sup>450</sup> Die vorgestellte Summationsproblematik bleibt damit allerdings erhalten.

DWH-spezifische Konzepte bieten damit – auch bei Berücksichtigung von Methoden des Ontologie-Matchings – bisher noch keinen umfassenden Ansatz, der (semi-)automatisch eine vollständige Integration verschiedener multidimensionaler Datenstrukturen ermöglicht. Es zeigt sich, dass eine Automatisierung umso einfacher wird, je mehr Veränderungs- und Neugestaltungsmöglichkeiten bei den Zielstrukturen zugelassen werden.<sup>451</sup> Je autonomer und unverbundener die ‚beteiligten Datenquellen‘ jedoch bleiben sollen, desto weniger können diese Möglichkeiten geschaffen werden. Im Folgenden werden daher Ansätze vorgestellt, die insbesondere die Unabhängigkeit der Teilnehmer berücksichtigen.

---

<sup>447</sup> Vgl. Abschnitt 4.3.1.

<sup>448</sup> Vgl. Riazati/Thom/Zhang (2011): 15.

<sup>449</sup> Vgl. Olaru (2012): 570.

<sup>450</sup> Vgl. Neumayr/Schrefl/Thalheim (2010): 61; Schütz et al. (2011): 25.

<sup>451</sup> Vgl. dazu insbesondere Olaru (2012); Torlone (2008).



#### 4.3.4 Ansätze mit besonderer Berücksichtigung unabhängiger Netzteilnehmer

Die vorhergehenden Betrachtungen haben gezeigt, dass eine Nähe zwischen PDMS und BI-Netzwerken besteht, wenngleich die Fragestellungen nicht identisch sind.<sup>452</sup> Weiterhin wurde gezeigt, dass die Ansätze einer GAV- oder LAV-Modellierung von Datenbanken nicht direkt auf den BI- oder DWH-Kontext übertragbar sind.<sup>453</sup> Es ergeben sich aber aus der Verwendung dieser Modellierungen im PDMS-Bereich Konzepte, die für die Gestaltung von P2P-basierten – und damit der Unabhängigkeit der Beteiligten entgegenkommenden – BI-Netzwerken adaptiert werden können. HALEVY ET AL. entwickeln ein virtuelles globales Schema, das mit GAV-Techniken abgefragt werden kann, dessen Materialisierung aber über einen LAV-Ansatz geschieht, sodass die Daten weiterhin ausschließlich lokal bei den Netzteilnehmern vorliegen.<sup>454</sup> Dabei ist es möglich, Abfragen mehrfach umzuformulieren, sodass kein vollständig verbundenes Netzwerk vorliegen muss.<sup>455</sup> Darauf aufbauend verwenden BONIFATI ET AL. für XML-basierte Datenbanken eine P2P-orientierte Implementierung, die eine formal ausführlich definierte Mapping-Sprache nutzt und eine visuelle Mapping-Hilfe besitzt.<sup>456</sup> Eine weitere Variante des Vorgehens stellen MCBRIEN/POULOVASSILIS vor, die den BAV-Ansatz nutzen, um Transformationspfade zwischen den Netzteilnehmern zu schaffen. Sie sehen allerdings keine mehrfache Umformulierung vor, um ein vollständig verbundenes Netzwerk als Voraussetzung zu eliminieren, sondern schlagen den Einsatz virtueller ‚Superpeer-Schemata‘ vor.<sup>457</sup> Dabei können einzelne Peers die Transformationspfade ihres lokalen Schemas zu einem (beliebigen) globalen Schema – dem Superpeer-Schema – öffentlich machen, sodass andere Teilnehmer durch die Schaffung eines Transformationspfades zum identischen Superpeer-Schema gleichzeitig eine Überführbarkeit zu allen verbundenen lokalen Schemata erzielen. Abfragen an das Netzwerk können dann zumindest an alle Peers verteilt werden, die zu mindestens einem der dem Abfragenden bekannten

---

<sup>452</sup> Vgl. Abschnitt 4.1.

<sup>453</sup> Vgl. Abschnitt 4.3.2.

<sup>454</sup> Vgl. Halevy et al. (2003): 506.

<sup>455</sup> Vgl. Halevy et al. (2003): 515. Die (hier) unerwünschte Eigenschaft eines globalen Schemas und die entsprechenden Einschränkungen bleiben aber wie bereits diskutiert erhalten.

<sup>456</sup> Vgl. Bonifati et al. (2010): 232–237.

<sup>457</sup> Vgl. McBrien/Poulovassilis (2004): 91f.

Superpeer-Schemata ebenfalls einen Transformationspfad definiert haben.<sup>458</sup> Eine Überführung der Superpeer-Schemata aufeinander oder das mehrfache Umformulieren von Abfragen über mehrere Peers und Superpeer-Schemata ist dabei nicht vorgesehen, sodass in einem P2P-Netzwerk zwar keine vollständige Verbundenheit gefordert wird, jedoch eine ‚Inselbildung‘ entstehen kann, wenn nicht mindestens ein Superpeer-Schema von allen Teilnehmern verwendet wird. Letzteres entspricht jedoch wieder der Idee eines föderierten DWH, deren Einschränkungen für die Autonomie der Teilnehmer diskutiert wurden.<sup>459</sup>

Um die Verwendung globaler Schemata zu vermeiden, schlagen VAISMAN/MINUTO ESPIL/PARADELA ein direktes Mapping der Datenbasen einzelner Peer-Paarungen vor, das auf Mapping-Tabellen basiert. Diese sind im P2P-Kontext erprobt und vorteilhaft anwendbar.<sup>460</sup> Dabei gehen die Autoren auf die DWH-Besonderheiten ein und überführen nicht nur die Schemata, sondern die Elemente,<sup>461</sup> was – wie in den bisherigen Überlegungen gezeigt – essenziell für die Aufrechterhaltung der Auswertbarkeit der Daten ist, insbesondere im Aggregationsfall. Anders als bei einem verhandelten globalen Schema steht der Abfragende Peer (im Weiteren zur Vereinfachung als ‚ $P_1$ ‘ bezeichnet) immer im Mittelpunkt. Seine Schema-Sicht soll bei Abfragen gewahrt bleiben, sodass das Mapping auch virtuelle Umordnungen (‚reclassify‘) und Aufteilungen (‚split‘) von Werten zulässt.<sup>462</sup> Die entsprechenden Umformulierungen sowie die dadurch geänderten Roll-Up-Funktionen werden auf dem Daten-liefernden Peer (‚ $P_2$ ‘) erzeugt und nur für Abfragen von  $P_1$  genutzt. Abb. 34 zeigt ein Beispiel für eine Umordnung. Element ‚3‘ wird bei  $P_1$  und  $P_2$  unterschiedlich eingeordnet, das Mapping schafft eine virtuelle Umordnung, die auf  $P_2$  gespeichert wird. Analog dazu stellt Abb. 35 einen Split dar, bei dem die zweite Aggregationsebene ein virtuelles neues Element erhält, was das Ziel des Roll-Ups von Element ‚4‘ darstellt und damit  $P_1$  zuordenbar ist.

---

<sup>458</sup> Vgl. McBrien/Poulovassilis (2004): 96f.

<sup>459</sup> Vgl. den einführenden Text zu Abschnitt 4.3.

<sup>460</sup> Vgl. Kementsietsidis/Arenas/Miller (2003): 325–327.

<sup>461</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 235f.

<sup>462</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 236.

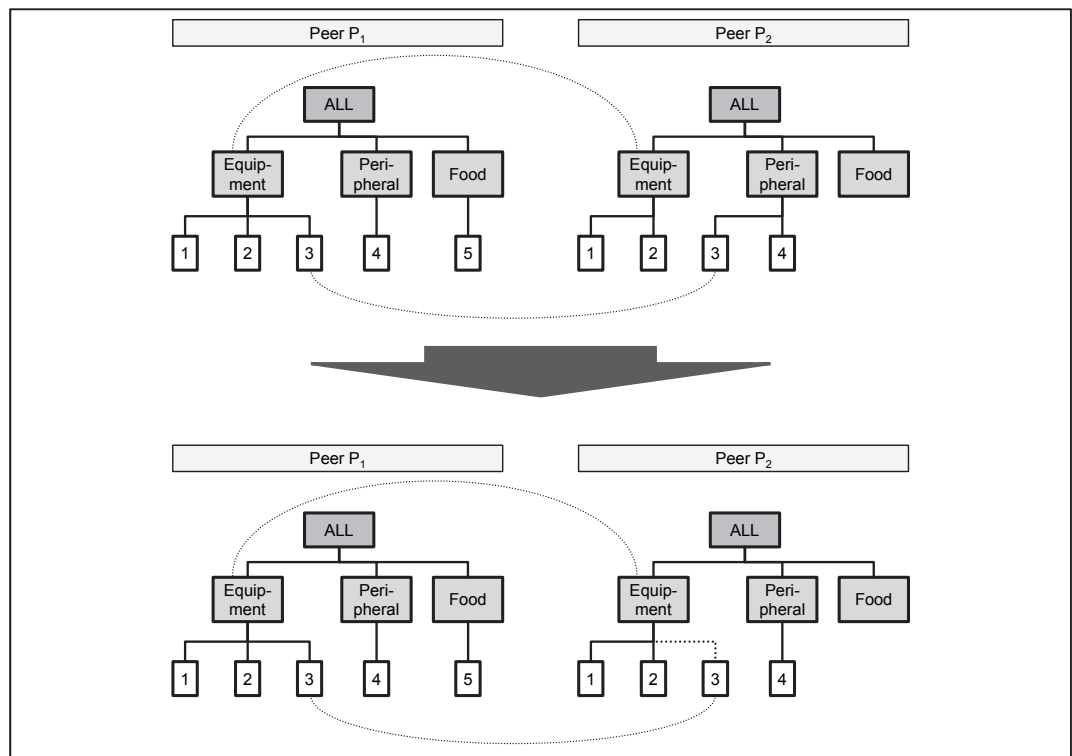


Abb. 34: Umordnung von Elementen einer ‚Produkt‘-Dimension

(Quelle: In Anlehnung an Vaisman/Minuto Espil/Paradela (2009): 237)

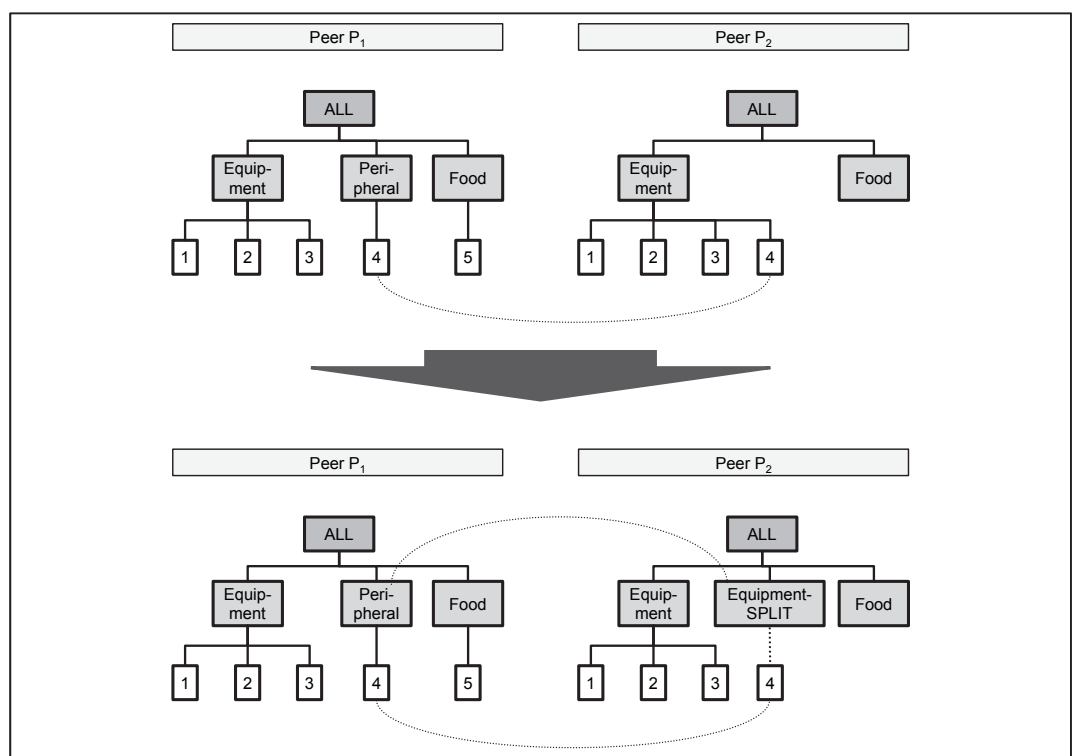


Abb. 35: Aufteilung von Elementen einer ‚Produkt‘-Dimension

(Quelle: In Anlehnung an Vaisman/Minuto Espil/Paradela (2009): 238)

Prinzipiell ermöglichen die Autoren damit eine Verkettung mehrerer P2P-Transformationen, schließen diese jedoch zunächst aus und beschreiben nur den Fall eines vollständig verknüpften Netzwerks.<sup>463</sup> Die Skalierbarkeit des Netzwerks ist damit gering, da bei dem Eintritt eines neuen Peers Verbindungen zu allen weiteren Peers aufgebaut werden müssen und dies zu einem quadratischen Anstieg der Verbindungen führt.<sup>464</sup>

GOLFARELLI ET AL. weisen zudem darauf hin, dass VAISMAN/MINUTO ESPIL/PARADELA nur den Austausch eines einzelnen Fakts betrachten und keine Inhomogenitäten der Dimensionselemente sowie unterschiedliche Datenformate berücksichtigen.<sup>465</sup> Der Ansatz des BIN<sup>466</sup> erweitert daher diese Überlegungen und ermöglicht zum einen eine Verkettung mehrerer Abfragen durch Aneinanderreihung mehrerer Peer-spezifischer Transformationen, zum anderen durch Annotationen an der Mapping-Tabelle auch die Überführung von weitergehender Element-Semantik.<sup>467</sup> Die Formulierung von Abfragen in einem BIN ist damit die weitgreifendste Erarbeitung einer Zusammenführung von unabhängigen Datenquellen und der entsprechenden Semantik. Allerdings blenden die Autoren wiederum die (teil-)automatische Generierung der Überführungen aus und gehen von einem manuell durchgeführten, zumindest aber fehlerfreien Mapping aus.

Alle vorgestellten Ansätze der vorhergehenden Abschnitte gehen daher – wie eingangs formuliert – auf einzelne, ggf. mehrere Aspekte der Harmonisierung von multidimensionalen Datenstrukturen im Rahmen von CCBI ein, ein umfassender Ansatz, der sowohl DWH-Spezifika als auch automatisierte Schema-Matching-Verfahren berücksichtigt, existiert derzeit nicht. Vereinzelt werden aber auch Aspekte genannt, die nicht direkt der Überführung, sondern vielmehr der Nutzung der Daten zuzuordnen sind, insbesondere hinsichtlich Sicherheits- und Vertraulichkeitsüberlegungen. Der folgende Abschnitt greift diese auf.

---

<sup>463</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 232.

<sup>464</sup> Es lässt sich direkt erkennen, dass für  $n$  Teilnehmer  $\binom{n}{2} = \frac{n^2-n}{2}$  Paarungen erzeugt werden müssen.

<sup>465</sup> Vgl. Golfarelli et al. (2012b): 395f.

<sup>466</sup> Vgl. die Ausführungen zum P2P-basierten ‚Business Intelligence Network‘ in Abschnitt 4.1 sowie die grafische Darstellung in Abb. 26.

<sup>467</sup> Beispielhaft wird der Fall genannt, dass bei der Überführung einer Dimension ‚Geschlecht‘ die Ausprägungen ‚M‘ und ‚F‘ in ‚Mann‘ und ‚Frau‘ überführt werden können (vgl. Golfarelli et al. (2012b): 399).

#### 4.4 Sicherheit und Vertrauen in Business-Intelligence-Netzwerken

Beim Aufbau von BI-Netzwerken soll – wie in anderen Fällen der Kooperation auch – gewährleistet sein, dass nur die gewünschte Menge an Daten mit den Kooperationspartnern geteilt wird. Zu jedem Zeitpunkt soll seitens jedes Teilnehmers bestimmt werden können, welche Daten vertraulich gehandhabt werden, d. h. Überlegungen zur Privacy<sup>468</sup> der Teilnehmer sind Bestandteil der Netzgestaltung, insbesondere in dezentral organisierten Netzwerken.<sup>469</sup> Für verteilte OLAP-Anwendungen können zwei Fragestellungen unterschieden werden, die sich zum einen auf die Gewährleistung der Nichtrückverfolgbarkeit von Einzeldaten, zum anderen auf die Sicherstellung der Weitergabe nur beabsichtigter Daten zur Erstellung von Aggregationen beziehen.<sup>470</sup> Wird ein zentraler OLAP-Würfel auf einem verteilten System erzeugt (ggf. nur temporär), so ist es für die Möglichkeit, individuelle Abfragen auf den Daten zu erzeugen, notwendig, die Detaildaten zu speichern und Aggregationen entsprechend der Abfragen zu berechnen. Damit diese Detaildaten nicht direkt verwendbar sind, können sie (global im Würfel oder lokal bei den bereitstellenden Systemen) pertubiert, also kontrolliert verändert werden unter der Aufrechterhaltung einer Aggregationsfähigkeit. Die Durchführung dieses Vorgehens, auch in P2P-Netzwerken, wurde von AGRAWAL/SRIKANT/THOMAS als ‚Privacy Preserving OLAP‘ erarbeitet.<sup>471</sup> Ein Framework für die Weiterleitung nur notwendiger (und optimalerweise auch verschlüsselter) Daten stammt von CUZZOCREA/BERTINO.<sup>472</sup> Beide Ansätze stellen für die Fragestellungen den aktuellen Forschungsstand dar. CUZZOCREA/BERTINO weisen aber auch darauf hin, dass beide Felder bisher sehr wenig untersucht wurden.<sup>473</sup>

Unabhängig von diesen auf OLAP bezogenen und Privacy-orientierten Überlegungen, sind Sicherheitsaspekte im Allgemeinen durch die Zunahme der Vernetzung im IT-Bereich von besonderer Bedeutung.<sup>474</sup> IT-Sicherheit, bzw. Informati-

---

<sup>468</sup> Der englische Begriff ‚privacy‘ kann am ehesten mit ‚Privatsphäre‘ übersetzt werden. Im Kontext von informationeller Selbstbestimmung (‚information privacy‘) und IT im Allgemeinen beschreibt dies jedoch nicht vollständig die Thematik. OCHS/LÖW verwenden den Begriff „Privatheit“ (Ochs/Löw (2012): 16), der im Deutschen eher unspezifisch wirkt. Wie bei anderen Begriffen auch, wird daher für die vorliegende Arbeit auf eine Übersetzung verzichtet.

<sup>469</sup> Vgl. Kementsietsidis/Arenas/Miller (2003): 325.

<sup>470</sup> Vgl. Cuzzocrea/Bertino (2014): 120.

<sup>471</sup> Vgl. Agrawal/Srikant/Thomas (2005): 251f.

<sup>472</sup> Vgl. Cuzzocrea/Bertino (2011): 965–969.

<sup>473</sup> Vgl. Cuzzocrea/Bertino (2014): 120.

<sup>474</sup> Vgl. Manshaei et al. (2013): 25:1f.

onssicherheit, wird dabei sowohl von privaten als auch von öffentlichen Stellen in diversen Facetten (physischer Zugriff, Systemzugänge, Risikoanalysen, etc.) betrachtet. Eine umfangreiche Übersicht zu allgemeinen und speziellen Konzepten sowie eigene Empfehlungen liefert z. B. das Bundesamt für Sicherheit in der Informationstechnik.<sup>475</sup> Die vorgestellten Überlegungen gelten dabei zunächst für nahezu jede Form von unternehmensbezogenem Einsatz von IT und lassen sich mit geringfügigen Anpassungen auf jede Form der (netzbasieren) IT-Unterstützung anwenden. Sie werden daher im Rahmen dieser Arbeit nicht detaillierter beschrieben.

Vor dem Hintergrund der Betrachtungen dezentraler und unabhängiger Netzteilnehmer erläutern KIRKMAN/DEZHGHOSHA Gefahren für P2P-Netzwerke sowie entsprechende Gegenmaßnahmen.<sup>476</sup> Diese lassen sich verschiedenen ‚Ebenen‘ zuordnen, zu denen das eigentliche, unterliegende Netzwerk gehört, die logische Schicht der verbundenen Teilnehmer sowie die Applikation mit den ausgetauschten Daten.<sup>477</sup> Nur die ‚höchste‘, Applikations-orientierte Schicht ist dabei für den BI-Kontext spezifisch zu sehen, sodass auch hier auf eine detaillierte Betrachtung der unteren, allgemein betrachteten Schichten verzichtet wird. Im Fokus stehen vielmehr Überlegungen zur Sicherstellung einer sinnvollen und mit dem Kooperationsgedanken zu vereinbarenden Nutzung eines CCBI-Netzwerks.

Zentral sind dabei Überlegungen zu Vertrauen („trust“), das die Netzteilnehmer einander gegenüber aufbringen müssen, um kooperativ zu handeln. Ein Informations-getriebenes Netzwerk kann bspw. seinen Zweck nicht erfüllen, wenn Informationen gefälscht oder vorenthalten werden und eine (größere) Menge an Mitgliedern eine Nicht-Kooperationsstrategie anwendet.<sup>478</sup> Vertrauen ist dabei die subjektiv wahrgenommene Wahrscheinlichkeit, mit der ein anderes Individuum die von ihm erwartete Aufgabe in einem vorgegebenen Rahmen tatsächlich er-

---

<sup>475</sup> Vgl. Bundesamt für Sicherheit in der Informationstechnik (2008): 8–12.

<sup>476</sup> Vgl. Kirkman/Dezhgosha (2011): 3–5.

<sup>477</sup> Vgl. Chen/Xu/Bruda (2010): 194. Eine ähnliche, ausführlichere Einteilung findet sich im ISO/OSI-Schichtenmodell zur Netzwerkprotokollbeschreibung wieder. Das Modell wird im Rahmen der Überlegungen zur Bildung eines Referenzmodells im Detail vorgestellt (vgl. Abschnitt 4.5.1).

<sup>478</sup> Die Überlegungen lassen sich aus dem Konzept der Spieltheorie ableiten, die bereits zur Rechtfertigung der Kooperation verwendet wurde (vgl. Abschnitt 2.1). Aus der Anwendung des Spieltheorie lassen sich daher für den kooperativen Rahmen ausführliche Sicherheitskonzepte entwickeln, die auch „free riders“, d. h. rein konsumierende – und damit unkooperative – Netzteilnehmer berücksichtigen (vgl. Manshaei et al. (2013): 25:12[sic!]).

füllt.<sup>479</sup> Bei der Nutzung von dezentralen Systemen mit mehreren Teilnehmern ergibt sich die Problematik der Vertrauensermittlung, -speicherung und -kommunikation. Unter der Annahme bestimmter Eigenschaften von Vertrauen – wie z. B. Transitivität<sup>480</sup> – lassen sich allerdings Metriken entwickeln, die einzelnen Teilnehmern nicht-manipulierbare Vertrauenswerte zuweisen.<sup>481</sup> Entsprechende Implementierungen liegen bereits seit mehreren Jahren vor und gewinnen auch im BI-Bereich an Relevanz.<sup>482</sup>

Die vorherigen Ausführungen beschreiben umfangreiche Themenfelder – sowohl im Allgemeinen als auch im Speziellen für CCBI. Mit Bezug auf die Ausführungen zum Fallbeispiel<sup>483</sup> und dem Fokus auf die fachliche Nutzbarkeit von CCBI-Netzwerken wird für die vorliegende Arbeit von einer kooperationsbereiten Menge an Individuen bzw. Unternehmen ausgegangen, die ‚fair‘ handelt und kein den allgemeinen Fall überschreitendes Sicherheitsrisiko erzeugt oder handhaben muss. Für die weitere Betrachtung wird daher eine Erfüllung der Notwendigkeiten für Privacy, Sicherheit und Vertrauen vereinfachend angenommen.

#### **4.5 CroCoBIN: Ein einfaches Referenzmodell für Peer-to-peer-Business-Intelligence-Netzwerke**

Die Vielzahl der Ansätze, die in den vorhergehenden Abschnitten vorgestellt wurden, zeigt, dass ein großer Teil der möglichen Aspekte eines unternehmensübergreifenden BI-Netzwerks bereits – wenn auch teilweise in anderem Kontext – angedacht oder sogar detailliert ausgearbeitet wurde. Es zeigt sich aber auch, dass kein vereinheitlichender Rahmen existiert, der sich die bestehenden Ansätze zunutze macht und für den in dieser Arbeit skizzierten Anwendungszweck zusammenführt.

Auch das vorgestellte BIN-Konzept berücksichtigt nur einen Ausschnitt der vorhandenen Ansätze, sodass sich offene Punkte ergeben, deren Behandlung in einem übergreifenden Modell vorgesehen werden kann – auch wenn eine Lösung ggf.

---

<sup>479</sup> Vgl. Gong/Zhu (2012): 116.

<sup>480</sup> Transitivität beschreibt den Effekt des Vertrauens in eine Person durch das existierende gemeinsame Vertrauen in einen Dritten (vgl. zu dem Effekt in P2P-Netzwerken Chen/Xu/Bruda (2010): 195.).

<sup>481</sup> Vgl. Gong/Zhu (2012): 117–119.

<sup>482</sup> Vgl. Ooi/Liau/Tan (2003): 4–6, 11; Chang/Dillon/Hussain (2006): 336, 339.

<sup>483</sup> Vgl. Abschnitt 2.4.

nicht Bestandteil des Modells ist. Dazu gehören die Fragen nach Optimierung der Laufzeiten – für die Caching-Ansätze in Frage kommen –<sup>484</sup> oder Sicherheitsprobleme.<sup>485</sup> Offen bleibt auch insbesondere die Frage nach der Messbarkeit einer Güte des instanziierten Modells.

Im Folgenden wird daher mit *CroCoBIN*<sup>486</sup> ein einfaches Referenzmodell für den Aufbau bzw. die Architektur von CCBI-Netzwerken vorgestellt, das die existierenden Ansätze aufgreift. Dabei wird die grundlegende Entscheidung zur Verwendung eines P2P-Ansatzes im Vorhinein getroffen. Dieser bietet vor dem Hintergrund unabhängiger Unternehmen, der Erkenntnisse zur grundsätzlichen Verwendbarkeit von P2P-Netzwerken im DWH-Kontext und der existierenden Arbeiten eine geeignete Basisstruktur. Auch in BI-verwandten Bereichen wie z. B. dem Wissensmanagement werden entsprechende Ansätze als vorteilhaft angesehen, wenn das Schaffen einer zentralen Verwaltungseinheit nicht möglich, ungeeignet oder unwirtschaftlich ist.<sup>487</sup> Weitere thematisch nahe Einsatzmöglichkeiten finden sich z. B. bei der Speicherung unstrukturierter Daten.<sup>488</sup>

Zur formalen Einordnung des Referenzmodells wird zunächst eine begriffliche Abgrenzung der entsprechenden Begriffe gegeben. Im Anschluss werden die Bestandteile des Referenzmodells erläutert und abschließend die bisher vorgestellten Ansätze zugeordnet sowie die Begründung für das weitere Vorgehen in der Arbeit gegeben.

#### 4.5.1 Abgrenzung des Referenzmodellbegriffs

Der Begriff des *Modells* – zunächst grob verstanden als strukturgleiche oder -ähnliche Abbildung eines Gegenstandsbereichs – ist in der Wirtschaftsinformatik nicht mit einer eindeutigen Definition belegt, sondern leitet sich aus allgemeinen Merkmalen und dem zugrunde gelegten wissenschaftstheoretischen

---

<sup>484</sup> Vgl. Abschnitt 4.2.2.

<sup>485</sup> Vgl. zu den Fragestellungen Abschnitt 4.1.

<sup>486</sup> *Cross-Company Business Intelligence Network*. Die Schaffung eines eigenen Namens dient in erster Linie der leichteren Referenzierung innerhalb der Arbeit. Der Einfachheit halber wird im Folgenden auch ein nach dem Referenzmodell gebildetes Netzwerk als CroCoBIN bezeichnet.

<sup>487</sup> Vgl. Maier/Hädrich (2006): 60. Die Autoren weisen dabei noch darauf hin, dass sich auch P2P-Ansätze wieder ausgehend von einem ‚reinen P2P‘ abschwächen oder modifizieren lassen, indem zentrale Knoten oder übergeordnete Knoten (‚superpeers‘, vgl. dazu auch Abschnitt 4.3.4) eingesetzt werden.

<sup>488</sup> Vgl. Cirani/Melegari/Veltri (2010): 1952. Die Abgrenzung dieser Einsatzgebiete, die im Kontext von ‚Big Data‘ genannt werden, wurde bereits in Abschnitt 4.2.1 diskutiert.



Verständnis ab.<sup>489</sup> Abbildungsorientiert beschreibt ein Modell eine reale Gegebenheit, konstruktionsorientiert hingegen eine mit Ersteller und Empfänger des Modells abgestimmte, den Grundsätzen der Vereinfachung und Pragmatik dennoch folgende Wahrnehmung der Gegebenheiten. Vor dem Hintergrund der stark systemisch geprägten Fragestellung eines CCBI-Netzwerks wird in dieser Arbeit ein abbildungsorientiertes Verständnis verwendet, wie es z. B. auch die *International Organization for Standardization* (ISO) im Bereich von Referenzarchitekturen verwendet: „[A] model [is] an abstract representation of reality in any form (including mathematical, physical, symbolic, graphical, or descriptive form) to present a certain aspect of that reality for answering the questions studied.“<sup>490</sup> Eingeschlossen sind dabei auch Modelle einer noch zu erreichenden Realität (‘Soll-Modelle’).<sup>491</sup>

In Abgrenzung dazu ist ein *Referenzmodell* ein Modell, das entweder zur Konstruktion weiterer Modelle verwendet wird oder mit der Intention, dies zukünftig zu tun, entwickelt wurde, dabei aber ebenfalls (vollständig oder teilweise) präskriptiv, d. h. zukunfts- oder vorschlagsorientiert gestaltet sein kann.<sup>492</sup> Diese weitgefaste Definition resultiert aus der Vielfalt von verschiedenen, als Referenzmodell bezeichneten Modellen oder Artefakten, zu denen auch wiederverwendbare Software-Bausteine oder Erkenntnissammlungen gehören können.<sup>493</sup> Klassifikationsansätze zu den verschiedenen Arten finden sich u. a. bei BROCKE und FETTKE/LOOS.<sup>494</sup> Entsprechend heterogen und situationsbezogen sind auch die Möglichkeiten zur Evaluation von Referenzmodellen beschrieben.<sup>495</sup> Bereits eine natürlich-sprachliche Evaluation wird, abhängig von Modell und Einsatzzweck, als sinnvoll angesehen, sollte aber, wenn möglich, mit anderen Methoden kombiniert werden.<sup>496</sup>

---

<sup>489</sup> Vgl. Strahringer (2013). Zu den wissenschaftstheoretischen Grundpositionen vgl. Abschnitt 1.2.1.

<sup>490</sup> ISO 15704 (2000): Abschnitt 3.16.

<sup>491</sup> Vgl. Strahringer (2013). Eine ausführliche Übersicht verschiedener Modelldefinitionen und -ausprägungen findet sich bei BURMESTER und den dort zitierten Autoren (vgl. Burmester (2011): 52–59).

<sup>492</sup> Vgl. Fettke/Brocke (2013).

<sup>493</sup> Vgl. Brocke (2003): 97.

<sup>494</sup> Vgl. Brocke (2003): 97–100; Fettke/Loos (2003): 43–48.

<sup>495</sup> Vgl. Gericke/Winter (2009): 203; Rosemann/Schütte (1997): 20f.

<sup>496</sup> Vgl. Fettke/Loos (2004): 7f., 20.

Eine Besonderheit der Referenzmodellierung stellen *Referenzarchitekturen* dar. So beschreiben z. B. BECKER/KNACKSTEDT die Architekturmodelle von DWH-Systemen<sup>497</sup> als Referenzmodell.<sup>498</sup> Eine *Architektur* ist (aus Sicht der Softwareentwicklung) die Struktur von Programm- und Systemteilen, deren Verbindungen sowie die zu Grunde liegenden Designrichtlinien. Im Gegensatz zu einem eher abstrakten Referenzmodell ist die Referenzarchitektur die Überführung des Modells auf Softwarekomponenten.<sup>499</sup> Wie uneinheitlich die Begrifflichkeiten sind, wird deutlich beim Einbezug weiterer, teilweise analog verwendeter Begriffe. So beschreiben NAKAGAWA/OLIVEIRA ANTONINO/BECKER ein Referenzmodell als abstraktes Framework, das Minimalübereinstimmungen von Modellen eines bestimmten Bereichs darstellt.<sup>500</sup> Der Framework-Begriff wird dabei nicht geklärt, an anderer Stelle aber als Werkzeug zur Architekturentwicklung definiert.<sup>501</sup> ANGELOV/GREFEN/GREEFHORST wiederum weisen darauf hin, dass auch eine softwarebezogene Referenzarchitektur nicht einem konkreten System oder auch nur dem Bedürfnis nach einem konkreten System entspringen muss, sondern die Vermutung einer zukünftigen Nützlichkeit bereits ausreichende Berechtigung für die Konstruktion ist.<sup>502</sup>

Im Rahmen dieser Arbeit wird CroCoBIN – basierend auf den vielfältigen Verständnismöglichkeiten – als Referenzmodell bezeichnet, das Elemente, Verbindungen und verschiedene Betrachtungsebenen eines CCBI-Netzwerks darstellt, Grundprinzipien zur Gestaltung des Netzwerks und seiner Anwendung festlegt und eine Instanziierung der Kernkomponenten ermöglicht. Das Referenzmodell folgt dabei, soweit möglich, den Grundsätzen ordnungsgemäßer Referenzmodellierung<sup>503</sup> und fokussiert entsprechend dem Grundsatz der *Relevanz* die Aspekte unternehmensübergreifender BI, die in bisherigen Arbeiten als besondere Schwierigkeit identifiziert wurden, d.h. die Harmonisierung der Datenbestände und die zu erhaltende Interpretierbarkeit derselben. Weiterhin wird der Grundsatz der *Wirtschaftlichkeit* berücksichtigt: CroCoBIN wird nicht mit dem Anspruch eines allumfassenden Modells erstellt, sondern dient der Erreichung des Erkenntnisziels

---

<sup>497</sup> Vgl. Abschnitt 4.2.1.

<sup>498</sup> Vgl. Becker/Knackstedt (2003): 417.

<sup>499</sup> Vgl. Bass/Clements/Kazman (2009): 24f.

<sup>500</sup> Vgl. Nakagawa/Oliveira Antonino/Becker (2011): 200.

<sup>501</sup> Vgl. Schönherr (2004): 16.

<sup>502</sup> Vgl. Angelov/Grefen/Greefhorst (2009): 141.

<sup>503</sup> Vgl. Rosemann/Schütte (1997): 17–25.

dieser Arbeit, wie und unter welchen Bedingungen P2P-CCBI-Netzwerke nutzungsfähig zu gestalten sind.<sup>504</sup> Entsprechend wird eine abstrakte Modellierung gewählt, die das Modell positiv in Bezug auf *Robustheit* und *Adaptierbarkeit* beeinflusst.<sup>505</sup>

#### 4.5.2 Aufbau des Referenzmodells

Der grundlegende Aufbau eines CroCoBIN, so wie in Abb. 36 dargestellt, ist direkt aus den bisher vorgestellten Überlegungen abzuleiten und ähnelt dem eines BIN.<sup>506</sup> Eine Anzahl  $n$  von Teilnehmern („Peers“) stellt Daten in einer multidimensionalen Struktur bereit und ist durch ein P2P-Netzwerk direkt oder indirekt miteinander verbunden. Jeder Teilnehmer verfügt über ein eigenständiges DWH, das mit einem – hier noch nicht näher spezifizierten – „Network Connector“, d. h. einem die Verbindung steuernden System, interagiert und so die Netzwerkfähigkeit sicherstellt.

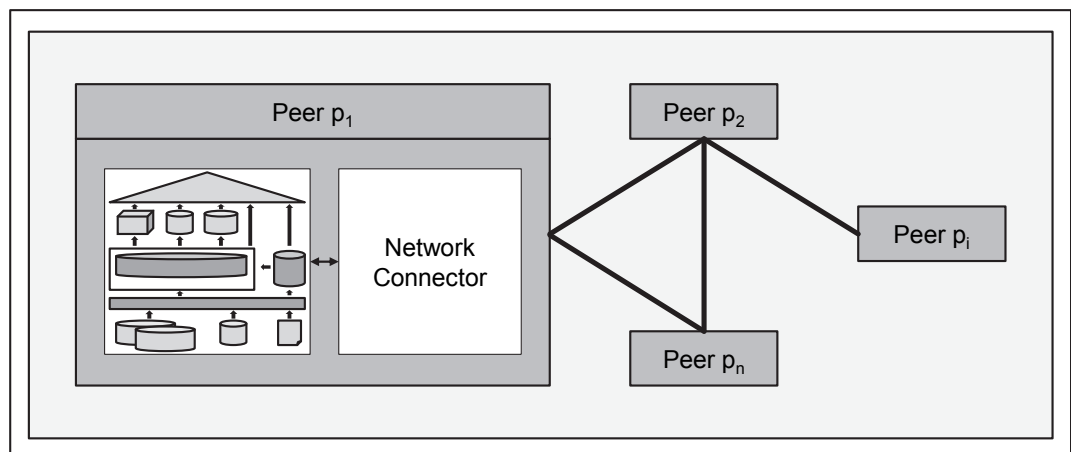


Abb. 36: Grundstruktur eines CroCoBIN

(Quelle: Eigene Darstellung)

Zur Beschreibung des Netzwerks und der Teilnehmer können mehrere Komponenten unterschieden werden. Da bisher kein umfassendes Modell oder Referenzmodell existiert, und im Kontext des BIN keine dedizierte Betrachtung aller

<sup>504</sup> Vgl. Abschnitt 1.1.

<sup>505</sup> Vgl. Rosemann/Schütte (1997): 21f. Weitere Grundsätze sind *Richtigkeit*, die laut den Autoren aber bei Referenzmodellen kaum prüfbar ist, *Klarheit*, die durch die Verwendung einfacher grafischer Strukturen, bekannter mathematischer Formulierungen und natürlich-sprachlicher Beschreibungen erreicht wird, *Vergleichbarkeit*, die erzielt wird, da z. B. das thematisch verwandte BIN-Modell in ähnlicher Form modelliert ist und semantisch verglichen werden kann sowie ein *systematischer Aufbau*, der durch den Bezug auf existierende und erprobte Modelle erreicht wird.

<sup>506</sup> Vgl. Abb. 26.

Netzbestandteile durchgeführt wird, wird zur Identifikation der einzelnen Komponenten auf thematisch nahe Konzepte zurückgegriffen. Neben dem allgemeinen Referenzmodell eines P2P-Netzwerks<sup>507</sup> ist dies insbesondere das Referenzmodell für semantische P2P-Netzwerke.<sup>508</sup> Ein SP2P-Netzwerk entspricht einem Netzwerk aus Peers, deren bereitgestellte Daten in Form einer Ontologie organisiert sind, sodass eine semantische Überleitung zwischen den Daten der Beteiligten geschaffen werden muss. Zur Entwicklung des SP2P-Referenzmodells betrachten MAWLOOD-YUNIS/WEISS/SANTORO existierende Implementierungen und leiten deskriptiv ein darauf anwendbares Modell ab. Dies führt zu einer Aufstellung diverser ‚Modell-Konzepte‘, die teilstrukturiert aufgenommen und generalisiert werden.<sup>509</sup> Sie beschreiben grundsätzliche Elemente des Modells, die sich der CroCoBIN-Struktur zuordnen lassen. Dazu wird diese zunächst in fünf Komponenten unterteilt, die sich aus der gezeigten Struktur und den Vorüberlegungen der Abschnitte 4.1 bis 4.4 ergeben (vgl. Abb. 37).

---

<sup>507</sup> Vgl. Aberer et al. (2005): 15–17.

<sup>508</sup> Vgl. Mawlood-Yunis/Weiss/Santoro (2011): 49–60. Die Autoren verwenden den Begriff der ‚Semantic Peer-to-Peer Networks‘ (SP2P).

<sup>509</sup> Vgl. Mawlood-Yunis/Weiss/Santoro (2011): 50.

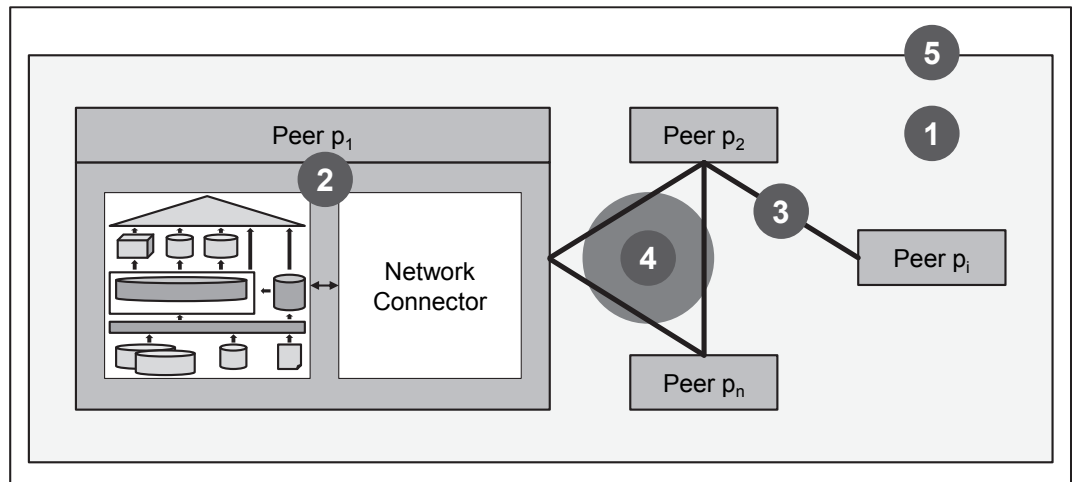


Abb. 37: Grundstruktur und Komponenten eines CroCoBIN

(Quelle: Eigene Darstellung)

- (1) Die *Netzwerkstruktur* beschreibt die grundsätzliche Verknüpfung der Teilnehmer des Netzwerks, die Überlegungen zur Implementierung und die notwendigen Netzwerkparameter.
- (2) Die *Peerstruktur* beschreibt den Aufbau eines Teilnehmers, bzw. Teilnehmer-Systems. Dazu gehört insbesondere die mögliche Ausgestaltung oder Ergänzung des 'Network Connector'.
- (3) *Prozesse einzelner Peer-Paarungen* beschreiben die direkte Interaktion zweier miteinander verbundener Peers, insbesondere die semantische Überleitung zueinander.
- (4) *Prozesse im Netzwerk* beschreiben das Gesamtverhalten des Netzwerks in Bezug auf Datenaustausch und Abfragen. Dazu gehört insbesondere das *Routing*, also die Entscheidung, über welche Teilnehmer indirekte Verbindungen hergestellt werden.
- (5) Die *organisatorischen Rahmenbedingungen* beschreiben die aus den übergreifenden Aspekten abgeleiteten Eigenschaften oder Voraussetzungen des Netzwerks, insbesondere hinsichtlich Sicherheit und Vertrauen.

Tab. 9 stellt die hier gewählten Komponenten den im SP2P-Referenzmodell erarbeiteten Modell-Konzepten gegenüber.

<b>CroCoBIN-Komponente</b>	<b>SP2P-Referenzmodell-Konzept</b>	
<b>Bezeichnung</b>	<b>Bezeichnung</b>	<b>Beschreibung</b>
<b>Netzwerkstruktur</b>	<b>Peer</b>	Netzwerkentität / Teilnehmer
	<b>Peer neighbors</b>	Referenzen eines Peers auf andere Peers im Netzwerk
	<b>Semantic neighborhood</b>	Referenzierte Peers, zu denen eine hohe semantische Ähnlichkeit besteht, sodass eine Überleitung möglich ist
	<b>Network degree</b>	Maximale Verbindungsanzahl eines Peers (vordefiniert)
<b>Peerstruktur</b>	<b>Peer profile</b>	Beschreibung eines Peers, seines bereitgestellten Inhalts und seines Informationsschemas
<b>Prozesse einzelner Peer-Paarungen</b>	<b>Mapping</b>	Semantische Überleitung zwischen den Ontologien einzelner Peers
	<b>Similarity function</b>	Funktion des Maßes der Stärke der semantischen Affinität zweier Peers
<b>Prozesse im Netz</b>	<b>Peer discovery</b>	Vorgehen von Peers zur Ermittlung semantisch naher Peers
	<b>Autonomous joining</b>	Entscheidung eines Peers über zu verknüpfende andere Peers
<b>Organisatorische Rahmenbedingungen</b>	<b>Keine Entsprechung</b>	

Tab. 9: Zuordnung der Modell-Konzepte des SP2P-Referenzmodells zu den Komponenten von CroCoBIN

(Quelle: Eigene Darstellung<sup>510</sup>)

Das SP2P-Referenzmodell und CroCoBIN weisen erkennbare Übereinstimmungen in den Komponenten auf, wobei das SP2P-Referenzmodell einzelne, weniger abstrakte Konzepte beschreibt, was auf die explorative Erstellung auf Basis existierender Modelle und die Identifikation von zentralen Bausteinen dieser zurückzuführen ist.<sup>511</sup> Die im Folgenden detailliert vorgestellten, für CroCoBIN verwendeten Komponenten greifen die entsprechenden Konzepte auf und erweitern sie bei Bedarf.

#### 4.5.2.1 Netzwerkstruktur

Die bisher in der Arbeit vorgestellten grundlegenden Konzepte von P2P-Netzwerken werden zunächst noch einmal zusammenfassend aufbereitet und formal beschrieben. Ein P2P-Netzwerk ist die Verbindung unterschiedlicher, aber

<sup>510</sup> Zur Aufzählung der SP2P-Modell-Konzepte vgl. Mawlood-Yunis/Weiss/Santoro (2011): 50.

<sup>511</sup> Vgl. Mawlood-Yunis/Weiss/Santoro (2011): 41.

gleichwertiger Teilnehmer über ein existierendes physisches Netzwerk. Jeder Teilnehmer ist dabei in der Regel sowohl Datensender als auch -empfänger, wobei dies auch ein reines Weiterleiten von Daten bedeuten kann.<sup>512</sup> Obwohl dies auch auf einen Großteil der Internet-Infrastruktur als solche zutrifft, wird als P2P-Netzwerk üblicherweise ein überlagerndes Netzwerk verstanden, das in die physische Struktur nicht eingreift. Das P2P-Netzwerk wird daher als ‚overlay network‘ bezeichnet.<sup>513</sup> Zur Unterscheidung der betroffenen Ebenen kann das Basismodell zur *Open Systems Interconnection* (OSI) verwendet werden, das die Netzinfrastruktur in sieben Schichten („layer“) unterteilt (vgl. Abb. 38).<sup>514</sup>

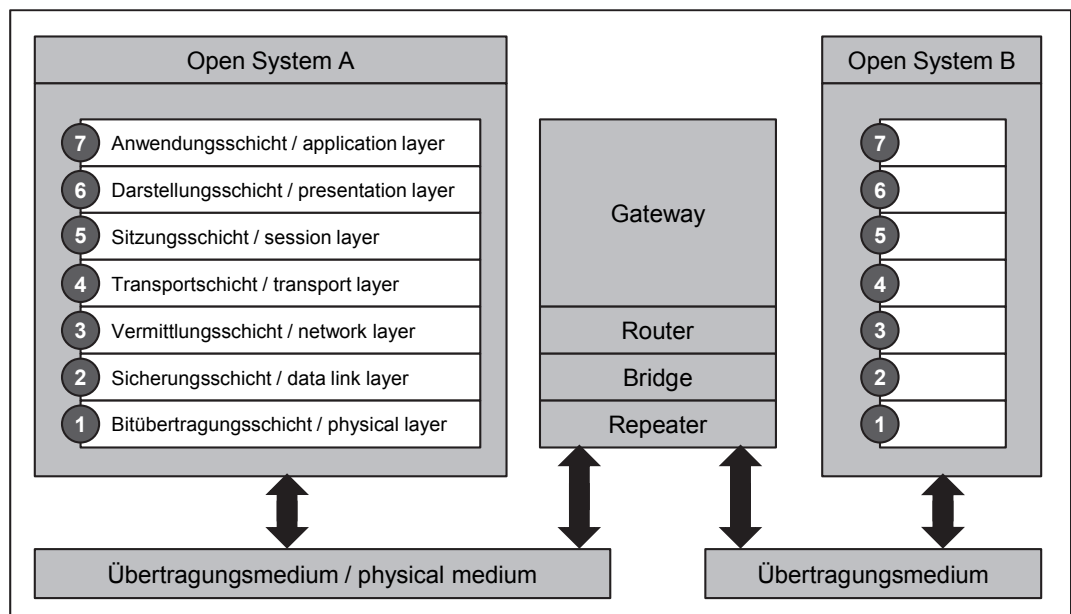


Abb. 38: OSI-Schichtenmodell

(Quelle: In Anlehnung an Keller (2011): 289f.)

Die Schichten eins bis vier dienen dabei dem reinen Transport von Daten(paketen) über ein Übertragungsmedium, während die Schichten fünf bis sieben anwendungsorientiert sind und die Prozesskommunikation (Sitzungsschicht), die Datenkonvertierung (Darstellungsschicht) sowie die eigentliche Datenein- und -ausgabe (Anwendungsschicht) steuern. Ein P2P-Netzwerk im beschriebenen Verständnis setzt ein funktionierendes Netzwerk der Schichten eins bis vier voraus und ‚überlagert‘ die dort (z. B. physisch) existierende Struktur mit einem

<sup>512</sup> Vgl. Abschnitt 4.1.

<sup>513</sup> Vgl. Aberer et al. (2005): 11; sowie die Ausführungen zum Begriff ‚Overlay‘ in Abschnitt 4.2.1.

<sup>514</sup> Vgl. International Telecommunication Union (1994).

logischen Netzwerk verbundener Teilnehmer.<sup>515</sup> Im Rahmen von CroCoBIN wird weiter vereinfachend ausschließlich die Anwendungsschicht betrachtet und ein funktionierendes Kommunikationsprotokoll zum Aufbau und Erhalt von Verbindungen ebenso wie eine Datenformat-homogenisierende Schicht vorausgesetzt, da die Betrachtung semantischer Verbindungen ausschließlich in der Anwendungsschicht erfolgt. Dies ergibt sich auch aus den identifizierten Unterschieden von P2P- zu SP2P-Netzwerken, die neben dem Punkt der Semantik die Ressourcenautonomie der Teilnehmer betonen. Anders als in P2P-Netzwerken, in denen das Teilen, Zwischenspeichern und Weiterverbreiten von Daten den Regelfall darstellt, ist in SP2P-Netzwerken – wie auch in CCBI-Netzwerken – eine vielfache Reproduktion der Daten nicht gewünscht und/oder möglich.<sup>516</sup>

In den grundlegenden Begriffen folgt CroCoBIN unabhängig von den Besonderheiten des Einsatzzwecks jedoch der P2P-Referenzarchitektur und daher auch der formalen Definition von ABERER ET AL.:

**Definition (12)** Die *Netzwerkstruktur eines CroCoBIN* wird gebildet durch eine Menge Peers  $P = \{p_1, \dots, p_i, \dots, p_n\}$ , die die Teilnehmer darstellt, eine Funktion  $F_P: P \rightarrow 2^P \setminus \emptyset$ , die jedem Peer mindestens einen verbundenen Peer zuordnet und damit die *Nachbarschaft* eines Peers  $N(p)$  definiert, und eine *Metrik*  $s: P \times P \rightarrow \mathbb{R}$ , die die Ähnlichkeit<sup>517</sup> zweier Peers abbildet.<sup>518</sup>

Ohne weitere formale Betrachtung gelten für ein CroCoBIN zudem folgende Annahmen über die Eigenschaften des Netzwerks:<sup>519</sup>

<sup>515</sup> Vgl. Aberer et al. (2005): 15f.

<sup>516</sup> Vgl. Mawlood-Yunis/Weiss/Santoro (2011): 40.

<sup>517</sup> ABERER ET AL. verwenden statt einer Ähnlichkeitsabbildung eine Distanzfunktion, (vgl. Aberer et al. (2005): 12) was natürlich-sprachlich betrachtet eine Umkehrung bedeutet, aus formaler Sicht aber erst mit der Definition und Interpretation der Funktion relevant wird.

<sup>518</sup> Ein so definiertes Netzwerk kann als gerichteter Graph  $G=(P, E)$  modelliert werden, in dem die Peers die Knotenmenge  $P$  bilden und  $E$  die Menge aller Verbindungen zwischen den Knoten darstellt, d. h.  $E = \bigcup_i N(p_i)$  (vgl. Aberer et al. (2005): 14). Dies ermöglicht im Weiteren die Betrachtung grundlegender Eigenschaften eines gerichteten Graphen zur Betrachtung des Netzes. Eine ausführliche Formulierung der Bestandteile des Netzes erfolgt im Rahmen der Betrachtung des Aufbaus von Netzwerken in Abschnitt 5.2.1.

<sup>519</sup> Die verwendeten Begriffe sind in Teilen der Graphentheorie (vgl. Abschnitt 5.2.1) entnommen oder dem Sprachgebrauch der bisherigen Ausführungen gemäß als natürlich-sprachliche Begriffe ohne formale Definition neu eingeführt.



- Die Verbindung zweier Peers meint, dass es eine – wie auch immer ausgestaltete – *semantische Überführbarkeit* von Daten der beiden Peers gibt. In Bezug auf das SP2P-Modell bedeutet dies, dass eine Nachbarschaft nur dann eine Nachbarschaft darstellt, wenn sie eine *semantische* Nachbarschaft ist, denn erst dann ist die Verbindung im Sinne eines CCBI-Netzwerks nutzbar.
- Jede Verbindung zwischen zwei Peers ist *eindeutig*, d. h. es gibt nur eine Verbindung zwischen zwei Peers, und teilsymmetrisch, d. h.  $p_1 \in N(p_2) \leftrightarrow p_2 \in N(p_1)$ , allerdings ohne den Anspruch einer gleichartigen oder ‚gleichwertigen‘ semantischen Überführung. Dies basiert auf der Annahme, dass das Mapping der Daten zweier Teilnehmer immer in gegenseitigem (Ein-)Verständnis erfolgt und daher eine Überführung beiderseits gewollt ist, jedoch aufgrund unterschiedlicher Aggregationsebenen ggf. ein unterschiedlich detailliertes Mapping erzeugt wird.
- Das Netzwerk ist nicht zwingend *vollständig* verbunden, d. h. es gibt nicht notwendigerweise für jedes Paar von Peers eine *direkte* Verbindung. Die Gründe für diese Forderung liegen in der Komplexität der Gegenannahme eines vollständig verbundenen Netzwerks.<sup>520</sup>
- Das Netzwerk ist *zusammenhängend*, d. h. es gibt für jedes Paar von Peers immer eine zumindest *indirekte* Verbindung. Andernfalls lägen aus fachlicher Sicht zwei Netzwerke vor, da nur eine Teilmenge der Knoten auf eine Abfrage antworten könnte. Eine maximale Anzahl an Verbindungen eines Peers ist hingegen nicht vorgegeben, d. h. der ‚network degree‘ des SP2P-Modells entspricht immer  $n-1$ .
- Das Netzwerk erlaubt *Zyklen*, d. h. indirekte Verbindungen eines Peers zu sich selbst, da sonst die Autonomie der Partnerwahl erheblich eingeschränkt wird und beim Ausfall einzelner Knoten sehr schnell die Forderung des Zusammenhangs gefährdet wird.<sup>521</sup>

---

<sup>520</sup> Vgl. Abschnitt 4.3.4.

<sup>521</sup> Azyklische Netzwerke werden im Kontext der effizienten Query-Reformulierung im Rahmen von PDMS diskutiert und sind – unabhängig von der schlechten praktischen Anwendbarkeit – ein wünschenswerter Fall (vgl. Halevy et al. (2004): 791; Halevy et al. (2003): 511). Es erklärt sich auch damit die nicht vollständige Übertragbarkeit von Ansätzen aus diesem Bereich (vgl. Abschnitt 4.1).

#### 4.5.2.2 Peerstruktur

Die Struktur eines einzelnen Peers ist in den bisher vorgestellten Ansätzen ähnlich. Ein Peer besteht zunächst aus einem Datenbestand und einem System, das diesen Datenbestand für das Netzwerk zugänglich macht.<sup>522</sup> Abb. 39 zeigt eine weiter ausdifferenzierte Form, die die Grundstruktur um einige aus den vorgestellten Ansätzen abgeleitete Funktionen erweitert.

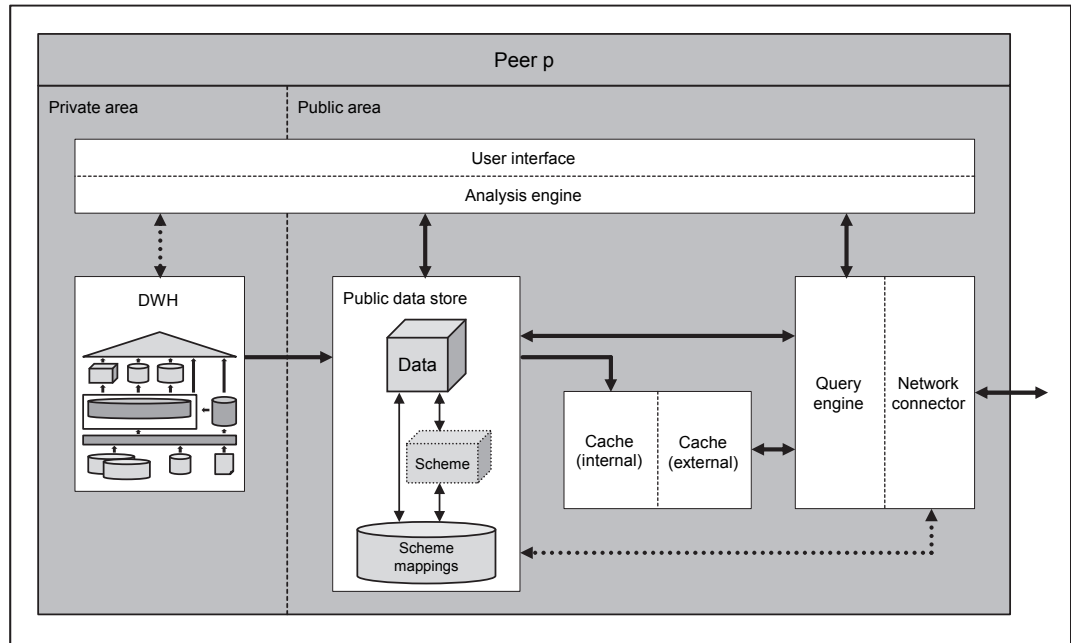


Abb. 39: Peerstruktur in einem CroCoBIN

(Quelle: Eigene Darstellung)

Zunächst bietet es sich aus Sicherheitsüberlegungen an,<sup>523</sup> die innere Struktur in einen internen („private area“) und einen öffentlich zugänglichen („public area“) Bereich zu unterteilen, wobei Daten, die über das Netzwerk verfügbar sein sollen, im öffentlichen Bereich liegen und entweder virtuell oder tatsächlich separat vorgehalten werden, sodass ein Zugriff auf weitere Daten nicht möglich ist. Der öffentliche Datenspeicherbereich („public data store“) enthält neben den eigentlichen Daten auch das zugrunde liegende Schema sowie eine Übersicht der Transformationen zu den Schemata der Peers in der Nachbarschaft. Ein Akteur stellt eine Anfrage, die mit Hilfe interner Daten und/oder einer Abfrage über das Netzwerk

<sup>522</sup> Vgl. den Ansatz des BIN (Abschnitt 4.1), die P2P-OLAP-Konzepte (Abschnitt 4.2.2) und den generellen Netzaufbau eines CroCoBIN (Einführung zu Abschnitt 4.5.2).

<sup>523</sup> Vgl. Abschnitt 4.4.

beantwortet wird (‘query engine‘ und ‘network connector‘).<sup>524</sup> Es kommen dabei als Anfragesteller sowohl Nutzer (über das ‘user interface‘) als auch automatisierte Analyseprogramme (‘analysis engine‘) in Frage.<sup>525</sup> Ebenfalls betrachtet werden die Überlegungen zu Caching-Verfahren im P2P-Kontext. Ein lokaler Cache hält sowohl Daten vor, die kürzlich selbst bereitgestellt wurden (‘internal cache‘), als auch solche, die kürzlich abgerufen wurden (‘external cache‘), unabhängig davon, ob dies zur eigenen Nutzung oder zur Weiterleitung auf Basis einer externen Anfrage geschah.<sup>526</sup>

Die vorgestellte Peer-Struktur berücksichtigt existierende Konzepte. Sie lässt zudem – durch Reduktion, Vereinfachen oder Auslassen von Bestandteilen – eine Implementierung weniger generischer oder umfangreicher Peer-Strukturen – wie in den vorgestellten Ansätzen (BIN, etc.) – zu und entspricht damit den Anforderungen an eine Referenzmodellierung.<sup>527</sup>

#### 4.5.2.3 Prozesse einzelner Peer-Paarungen

Notwendige Voraussetzung für den Datenaustausch zwischen zwei Peers ist die Überleitung der Daten in die jeweiligen Schemata. Die in Abschnitt 4.3 vorgestellten Matching-Verfahren zeigen, dass für den P2P-Fall ein Mapping zwischen je zwei Beteiligten, d. h. eine Zuordnung der Schema-Elemente und Daten-Elemente, die umsetzungsfähigste Variante darstellt – insbesondere im Vergleich zu globalen Schemata. Die Komplexität einer solchen Transformation, insbeson-

---

<sup>524</sup> Die Verwendung interner Daten stellt keine über das Netzwerk direkt zugängliche Möglichkeit dar und wird deshalb mit einer gestrichelten Linie symbolisiert. Gleiches gilt für den Austausch von Schemata über das Netz, für die ein separater Weg zu definieren ist, der hier nicht im Fokus steht.

<sup>525</sup> Die Verwendung automatisierter Analyseprogramme, die ein Netzwerk selbstständig nutzen können, erweitert die reine Datenbereitstellung zur ad-hoc-Analyse, wie sie z. B. im Rahmen des BIN beschrieben wird, um verteilte Rechenverfahren, wie sie CCBI grundsätzlich ermöglichen soll (vgl. Definition (6)) und die im Rahmen von CBI-PA als Abgrenzung zur reinen Datenbereitstellung beschrieben wurden (vgl. Abschnitt 3.3.3.3). Überlegungen zu verteilten Analyseformen finden sich bereits im Kontext verteilter Data-Mining-Verfahren, die hier nicht weiter thematisiert werden sollen. Eine Einführung in die und Übersicht der Thematik bieten ZENG ET AL. (vgl. Zeng et al. (2012)).

<sup>526</sup> Vgl. zu den Caching-Verfahren Abschnitt 4.2.2. Eine weitergehende Betrachtung der unterschiedlichen, in den entsprechenden Verfahren vorgestellten Caching-Strategien und Optimierungen (z. B. ‘pre-fetching‘ / ‘pre-caching‘) wird hier nicht vorgenommen. Das Vorhalten eines Caches widerspricht aufgrund seiner temporären Natur nicht der formulierten Annahme einer nicht-gewollten Ressourcen-Multiplikation (vgl. Abschnitt 4.5.2.1).

<sup>527</sup> Vgl. Abschnitt 4.5.1. Nicht berücksichtigt werden hier unterschiedliche systemische Voraussetzungen und die dafür ggf. notwendigen Transformationen oder Formatüberführungen, da dies zum einen von der tatsächlichen Instanz abhängig ist, zum anderen in der vorliegenden Arbeit eine fachliche und anwendungsorientierte Perspektive eingenommen wird (vgl. Abschnitte 4.5.1 und 4.5.2.1).

dere unter der Bedingung, dass eine erneute Reformulierung in ein weiteres Schema ermöglicht wird, wurde in Abschnitt 4.3.4 gesondert diskutiert. CroCoBIN verlangt daher weder eine spezifische Implementierung des Mappings, noch des eingesetzten Matching-Verfahrens. Da in der Kommunikation zweier Peers jeder Peer die Überleitungsregeln vorhält und damit in der Lage ist, die Abfrage des jeweils anderen Peers zu beantworten und weiterhin unabhängig von der Abfrage ausschließlich eins-zu-eins-Kommunikation erfolgt, ist es je nach Ausgestaltung der Query-Engine und der internen Prozesse des public data store auch möglich, paarweise individuelle Mappings und Matching-Verfahren einzusetzen. Dies kann u. a. dann hilfreich sein, wenn semantisch ähnliche Teilnehmer (z. B. Tochtergesellschaften eines Konzerns oder Mitglieder einer Produktionsgenossenschaft) mit semantisch weniger ähnlichen Teilnehmern (z. B. Gesellschaften eines anderen Konzerns) kooperieren und für das Matching der semantisch ähnlichen Peers (semi-)automatische, für das Matching zu weiteren Peers aber manuelle Verfahren eingesetzt werden sollen oder bei organisatorisch nahstehenden Peers ein (teil-)einheitliches Schema vereinbart wird.<sup>528</sup>

Sofern ein Mapping zwischen zwei Peers existiert, können Daten ausgetauscht werden. Das Vorgehen wird dabei in weiten Teilen durch die Struktur der Peers, wie im vorherigen Abschnitt beschrieben, bestimmt. Der anfragende Peer sendet die Schemainformation der gewünschten Fakten an den empfangenden Peer, der diese in sein lokales Schema überführt und die so zugeordneten Daten im Schema des anfragenden Peers zurückgibt. Für den Fall einer nicht vollständigen Überführbarkeit der Daten, die – z. B., weil der angefragte Peer keine Daten der gewünschten Granularität vorhält – zu einer ergebnislosen oder unvollständigen Beantwortung der Anfrage führt, ist netzübergreifend zu definieren, wie dies dem Empfänger der Daten angezeigt wird. Auch beim Einsatz des von VAISMAN/MINUTO ESPIL/PARADELA vorgeschlagenen „bottom-up-completing“<sup>529</sup>, das in jedem Fall eine Rückgabe von Daten ermöglicht,<sup>530</sup> muss dem Empfänger der Daten aufgezeigt werden, dass abgeleitete Daten und nicht originär granulare Da-

---

<sup>528</sup> Das Konzept von übergeordneten Schemata wurde im Rahmen von ‚superpeers‘ diskutiert (vgl. Abschnitt 4.3.4). Im Sinne eines generalisierenden Ansatzes besteht keine Notwendigkeit, die Möglichkeiten zur Vereinheitlichung einzuschränken, solange über die Gesamtheit des Netzwerks die Forderung nach möglicher Unabhängigkeit der Peers und ihrer Schemata aufrechterhalten wird.

<sup>529</sup> Vgl. Vaisman/Minuto Espil/Paradela (2009): 233.

<sup>530</sup> Vgl. dazu die ausführliche Darstellung des entsprechenden Matching-Vorgehens in Abschnitt 4.3.4.

ten zur Verfügung gestellt werden. Zur Darstellung der Güte der Überführung sind daher ein oder mehrere Maßzahlen (engl.: ‚measures‘) notwendig, die aus der Ähnlichkeit der Schemata und der Vollständigkeit des Mappings resultieren.<sup>531</sup>

#### 4.5.2.4 Prozesse im Netz

Die Messung der Überführungsqualität wird insbesondere bei der mehrfachen Überführung von Daten notwendig, wie sie bei indirekten Verbindungen in einem CroCoBIN regelmäßig auftritt. Sobald für eine indirekte Verbindung zwischen zwei Peers unterschiedliche Wege existieren, besteht bei unterschiedlichen Überführungen der Peers auf dem Weg die Möglichkeit, dass je nach Abfrageweg unterschiedliche Ergebnisse geliefert werden, die entsprechend unterschiedlich gut geeignet sind, die Abfrage bzw. die dahinterliegende Fragestellung, zu beantworten. Für das Routing der Abfragen in einem Netzwerk ist im allgemeinen Fall die Minimierung der Kosten, d. h. die Minimierung der Distanzfunktion, die leitende Entscheidung.<sup>532</sup> In einem CroCoBIN entspricht diese Distanzfunktion der Metrik  $s$ , die aufgrund ihrer Definition als Ähnlichkeitsabbildung maximiert werden muss.<sup>533</sup> Auf dieser Basis wird zur Vermeidung von Aggregationsfehlern<sup>534</sup> genau ein Weg ausgewählt, der sowohl für die Abfrage als auch für die Antwort verwendet wird. Die möglichen Wege werden dabei durch die Nachbarschaft jedes Peers bestimmt. Obwohl für den SP2P-Fall der Prozess des ‚peer discovery‘, also der Feststellung semantischer Übereinstimmungen zu anderen Teilnehmern, beschrieben wird, ist in allen vorgestellten Ansätzen diese Nachbarschaft exogen bestimmt, sodass hier kein Referenzvorgehen zur Nachbarschaftswahl diskutiert wird.<sup>535</sup>

Zur Vermeidung zyklischer Abfrageweiterleitungen und nicht abgeschlossener Abfragen durch Ausfall von Verbindungen, setzen P2P-Netzwerke Verfahren ein, die das mehrfache Bearbeiten der gleichen Abfrage erkennen und verhindern

---

<sup>531</sup> Ein entsprechendes Konzept wurde im SP2P-Kontext als ‚similarity measure‘ eingeführt (vgl. Tab. 9).

<sup>532</sup> Vgl. Aberer et al. (2005): 15; Abschnitt 4.2.2.

<sup>533</sup> Vgl. Definition (12).

<sup>534</sup> Vgl. Abschnitt 4.3.1.

<sup>535</sup> Die Erarbeitung von Empfehlungen zur Nachbarschaftswahl und der dafür notwendigen Measures entspricht einem wesentlichen Teil des Erkenntnisziels dieser Arbeit und wird in den Kapiteln 5 und 6 ausführlich diskutiert und validiert. In SP2P-Netzwerken wird von einer automatischen Ermittlichkeit der Nähe ausgegangen, was aufgrund der komplexen Überführungen in DWH-Umgebungen ex ante nicht praktikabel ist, wie die Ausführungen zu Überführungsverfahren in Abschnitt 4.3.3 zeigen.

(„cycle prevention policies“<sup>536</sup>) sowie die Lebensdauer der Abfragen begrenzen („time-to-live“<sup>537</sup>). Da diese Betrachtungen auf Ebene der Sitzungsschicht liegen,<sup>538</sup> wird für CroCoBIN kein spezifisches Vorgehen zum reinen Datenrouting festgelegt, sondern der konkreten Implementierung überlassen.

#### 4.5.2.5 Organisatorische Rahmenbedingungen

Anders als Netz- oder Peer-Spezifika sind die organisatorischen Rahmenbedingungen von CroCoBIN für jede Implementierung individuell von den Teilnehmern entsprechend der Bedürfnisse festzulegen. Die folgende Aufzählung stellt (mögliche) Kernelemente vor, wie sie in den vorgestellten Ansätzen und bei den Überlegungen zu Sicherheit und Vertrauen aufgebracht werden.<sup>539</sup>

- Mindestanforderungen der Teilnehmer in Bezug auf bereitgestellte *Hardware und Netzanbindung* sollen eine adäquate Beantwortung der Abfragen ermöglichen.
- Eine höhere Mindestgröße der Nachbarschaft ( $\geq 2$ ) kann zu einer höheren Zahl möglicher Überleitungswege führen, was sowohl die *Wahrscheinlichkeit einer guten Überleitung* der Schemata als auch die *Ausfallsicherheit* des Netzwerks erhöht.
- Zur Erhöhung der *Datensicherheit* kann eine Verschlüsselung der Verbindungen im Netzwerk definiert werden.<sup>540</sup>
- Unter Berücksichtigung von *Privacy-Aspekten* und unter Abwägung der *Nutzbarkeit des Netzwerks* kann ein Mindestumfang an freizugebenden Daten spezifiziert werden (im Anwendungsbeispiel könnte dies z. B. ein Mindestzeitraum vergangener Produktentwicklungen sein).
- Zur Sicherstellung des *Vertrauens* aller Beteiligten in das Netzwerk können Kontrollen installiert werden, die regelmäßig überprüfen, ob alle Beteiligten die gestellten Anforderungen erfüllen – und keine reinen ‚Datensammler‘ (‚free rider‘) existieren.

<sup>536</sup> Mawlood-Yunis/Weiss/Santoro (2011): 43.

<sup>537</sup> Ebd.

<sup>538</sup> Vgl. die Ausführungen zum OSI-Modell in Abschnitt 4.5.2.1.

<sup>539</sup> Vgl. Abschnitte 4.1 bis 4.4.

<sup>540</sup> Übliche Verschlüsselungsverfahren sind Transport Layer Security (TLS) oder der Vorgänger Secure Socket Layer (SSL). Beide sichern eine Verbindung auf der Transportschicht des OSI-Schichtenmodells (vgl. Saito/Sekiguchi/Hatsugai (2008): 252) und liegen somit nicht im direkten Definitionsbereich von CroCoBIN. Innerhalb des Modells kann aber die Verwendung bestimmter zugrunde liegender Netzwerkstrukturen beschlossen werden.

### 4.5.3 Gesamtmodell mit Einordnung bestehender Ansätze und Ableitung notwendiger Entwicklungen

Abb. 40 stellt abschließend die vorgestellten Komponenten als Gesamtmodell dar. Es werden dabei einheitlich die englisch-sprachigen Begriffe verwendet, um ein homogenes Bild zu erzeugen.

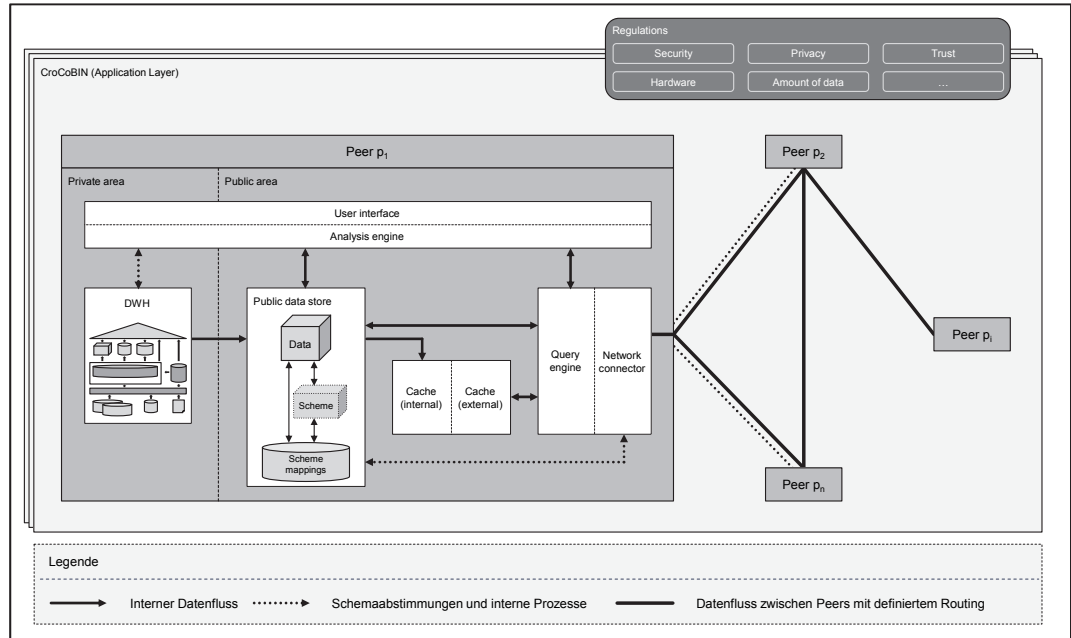


Abb. 40: CroCoBIN-Referenzmodell (Gesamtdarstellung)

(Quelle: Eigene Darstellung)

Für das vorgestellte Referenzmodell können auf Basis der Komponentenbeschreibungen und der zugehörigen Vorarbeiten im SP2P-Bereich (vgl. Tab. 9) die existierenden Ansätze hinsichtlich ihrer Eignung zur Ausgestaltung der einzelnen Komponenten zugeordnet werden. Tab. 10 stellt für die in diesem Kapitel diskutierten Ansätze dar, welche zentralen Aspekte der Komponenten von CroCoBIN in diesem Sinne behandelt oder angesprochen werden.<sup>541</sup>

<sup>541</sup> Zuordnungen werden durch eine Graufärbung markiert. Sofern ein Aspekt zwar thematisiert wird, aber kein Konzept / keine Lösung / kein Umsetzungsvorschlag genannt wird, erfolgt die Markierung mit einer 50%-Graufärbung. Dies trifft auch auf methodisch ausgearbeitete, thematisch jedoch ungeeignete Ansätze zu, wie z. B. die linguistischen Ähnlichkeitsmessungen zur automatischen Schemazusammenführung (vgl. Abschnitt 4.3.3), die nicht zur Messung der Ähnlichkeit zweier (bereits ineinander überführter) Schemata im fachlichen Sinne geeignet sind.

	Netz-Struktur		Peer-Struktur		Peer-Paarungen		Prozesse im Netz		Rahmenbedingungen	
Referenz	Allgemeiner Aufbau	Netzwerkparameter	Allgemeiner Aufbau	Caching	Matching / Mapping	Ähnlichkeitsmaß	Routing	Nachbarschaftswahl	Sicherheit / Privacy	Nutzbarkeitsvorgaben
Agrawal/Srikant/Thomas (2005)										
Akinde et al. (2003)										
Albrecht/Lehner (1998)										
Aouiche/Jouve/Darmont (2006)										
Banek et al. (2007)										
Bergamaschi et al. (2011)										
Berger (2009)										
Bernstein et al. (2005)										
Cali et al. (2003)										
Cuzzocrea/Bertino (2014) / Cuzzocrea/Bertino (2011) /										
Doan et al. (2004)										
Doka/Tsoumakos/Koziris (2011)										
Golfarelli et al. (2012b)										
Gong/Zhu (2012)										
Halevy (2001)										
Ji/Haase/Qi (2011)										
Kalnis et al. (2002)										
Kehua/Manirakiza (2012)										
Kirkman/Dezhgosha (2011)										
McBrien/Poulovassilis (2003)										
McBrien/Poulovassilis (2004)										
Neumayr/Schrefl/Thalheim (2010)/ Schütz et al. (2011)										
Olaru (2012)										
Rahm/Bernstein (2001)										
Riazati/Thom/Zhang (2011)										
Sapia (2000)										
Seshadri/Cooper/Liu (2005)										
Shvaiko/Euzenat (2013) / Shvaiko/Euzenat (2005)										
Torlone (2008)										
Vaisman/Minuto Espil/Paradela (2009)										
Wehrle/Miquel/Tchounikine (2005)										
Wu et al. (2009)										

Tab. 10: Zuordnung vorgestellter Ansätze zu Komponenten und zentralen Aspekten des CroCoBIN-Referenzmodells

(Quelle: Eigene Darstellung)



Deutlich erkennbar ist die Ungleichverteilung der Ansätze hinsichtlich der Behandlung der einzelnen Komponenten und Aspekte. Nur in geringem Maße wurden im Rahmen der Matching-Verfahren auch Ähnlichkeitsmaße untersucht, die sich dann in der Regel nicht auf multidimensionale Datenstrukturen und/oder -schemata beziehen, sondern auf einzelne Teile von Ontologien. Ebenfalls unbeachtet bleiben bisher die aktive Ausgestaltung von Nachbarschaften zur Wahl eines geeigneten Routing-Pfades und die damit verbundenen Empfehlungen zur Netzgestaltung anhand von Parametern oder Vorgaben – wobei letztere, wie in Abschnitt 4.5.2.5 formuliert, von der jeweiligen Ausprägung eines konkreten Netzwerks abhängen und in der vorliegenden Arbeit nicht thematisiert werden.

Damit stellt sich eine erhebliche Lücke in den Bereichen der Ähnlichkeitsmessung (zwischen Peer-Schemata im Ganzen) und des informationsmaximierenden Routings dar. Beide Bereiche sind jedoch für die Nutzbarkeit von CCBI-Netzwerken von entscheidender Bedeutung. So formuliert MELKAS für den zunehmenden Gebrauch von unternehmensübergreifenden Netzwerken: „[...] information-related issues require particularly urgent attention. [...] Challenges appear to be especially numerous there. Tools for analyzing information quality in such environments and on the basis of qualitative data have been lacking.“<sup>542</sup> Diese Aussage entspricht den bereits aufgegriffenen, themenspezifischen Einschätzungen von RIZZI zum Routing und zu den einzusetzenden Ähnlichkeitsmaßen, von TORLONE sowie RIAZATI/THOM/ZHANG zur Übereinstimmung von Schemata und schließlich von GOLFARELLI ET AL. zum Forschungsbedarf im Bereich Routing und Nachbarschaftswahl.<sup>543</sup> Im Folgenden werden daher diese Aspekte in den Vordergrund gestellt und auf Basis des entwickelten Referenzmodells betrachtet.

---

<sup>542</sup> Melkas (2004): 74f.

<sup>543</sup> Vgl. Golfarelli et al. (2012b): 410; Riazati/Thom/Zhang (2011): 9f.; Rizzi (2012): 201–203; Torlone (2008): 70–73.

## 5 Sicherstellung der Nutzbarkeit von Business-Intelligence-Netzwerken

Die bisherigen Betrachtungen haben gezeigt, dass die Nutzbarkeit von CCBI-Netzwerken – im Sinne des Nutzens, den jeder Teilnehmer aus der Teilnahme an der Netzgemeinschaft ziehen kann – nicht direkt von den einzelnen Teilnehmern, sondern von deren Interaktion abhängt. Eine Analyse von CCBI-Netzwerken, ihrer charakteristischen Eigenschaften und ihrer Gestaltungsmöglichkeiten muss daher auf einer Gesamtnetzebene stattfinden, wie sie auch die sozialwissenschaftliche Forschung abgrenzt von der (deutlich häufiger durchgeführten) Betrachtung einzelner Individuen in einem Netzwerk. Die Sicherstellung einer allgemeinen Nutzbarkeit oder Güte des Netzwerks kann daher dazu führen, dass lokale Optima – d. h. für CCBI-Netzwerke: Bestmögliche Abfrageergebnisse für ein teilnehmendes Unternehmen – ggf. nicht erreicht werden, um im Gegenzug der Gesamtheit der Teilnehmer ein gutes Ergebnis zu bieten.<sup>544</sup> PROVAN/FISH/SYDOW bieten eine 20 Jahre umfassende Übersicht der sozial- und betriebswirtschaftlichen, empirisch gestützten Forschung zu unternehmensübergreifenden Netzwerken auf Gesamtebene. Sie identifizieren drei Hauptforschungsstränge im Bereich *Netzwerkeigenschaften*: Netzwerkstrukturen, Netzwerkentwicklungen und Netzwerkerhaltung.<sup>545</sup> Relevant vor dem Hintergrund der ‚üblichen‘ Teilnehmer eines CCBI-Netzwerks sind dabei u. a. die Betrachtungen zur Netzwerkbildung unter Berücksichtigung von bereits existierenden Partnerschaften der Teilnehmer. Diese beziehen bei der Partnerwahl im Netzwerk auch die bisherigen Erfahrungen ein, bilden im Nachhinein aber nicht notwendigerweise auch stabile Partnerschaften.<sup>546</sup> Für eine unvoreingenommene, begründete Auswahl der Partner sollten daher in einem CCBI-Netzwerk entsprechend geeignete Strategien verwendet werden.<sup>547</sup>

---

<sup>544</sup> Vgl. Provan/Fish/Sydow (2007): 480, 483–485.

<sup>545</sup> Vgl. Provan/Fish/Sydow (2007): 502–505.

<sup>546</sup> Vgl. Provan/Fish/Sydow (2007): 503. Die Wahl von Partnern in neuen Netzen basierend auf bestehenden Bekanntschaften kann auch in Einzelpersonen-basierten virtuellen Umgebungen beobachtet werden, in denen die räumliche Nähe der Partner einen größeren Einfluss auf die Partnerschaft hat als andere Faktoren (vgl. Huang et al. (2009): 359). Allgemein wird das Phänomen in sozialen Netzwerken als *Homophilie* (von griech.: Gleiches liebend) beschrieben und führt zu der Bildung von Teilnetzen, die sich an bereits existierenden Eigenschaften der Teilnetzpartner ausrichten (vgl. Kim/Altmann/Hwang (2011): 3f.; McPherson/Smith-Lovin/Cook (2001): 415–417).

<sup>547</sup> Vgl. Abschnitt 5.4.

Die identifizierten *Netzwerkergebnisse*, die die Nutzbarkeit eines Netzwerks beschreiben, unterteilen sich in Performance- und Effektivitätsbetrachtungen, wobei Definition und Messung von Effektivität in diesem Kontext nicht eindeutig sind.<sup>548</sup> Im Bereich der P2P-Netzwerke weisen Performance-Betrachtungen hingegen klare Messgrößen auf. So können Abfragezeiten und Transfergrößen verhältnismäßig einfach bestimmt und Einflüsse von Netzwerkänderungen auf diese identifiziert werden.<sup>549</sup> In größeren Netzwerken lassen sich ‚bessere‘ Werte z. B. durch eine geschickte Auswahl der Partner (‚Nachbarschaft‘) erreichen, falls dadurch die Bildung von Teilnetzwerken, die bereits alle benötigten Informationen enthalten, ermöglicht wird,<sup>550</sup> was im Falle redundant vorgehaltener Daten zwar relevant, für den CCBI-Fall aber nicht anwendbar ist.<sup>551</sup> Weitere erhebliche Unterschiede in Performance-Fragen für ‚klassische‘ P2P-Netzwerke (z. B. für Filesharing<sup>552</sup>) und CCBI-Netzwerke ergeben sich aus der Netzgröße, die bei CCBI-Netzwerken ganz erheblich unter der von Filesharing-P2P-Netzwerken liegt.<sup>553</sup> Auch wissenschaftliche Publikationen thematisieren Performance-Fragen in Größenordnungen, die die wenigen Dutzend bis Hundert Nutzer von CCBI-Netzwerken erkennbar überschreiten.<sup>554</sup> Zudem ist die Abfragegeschwindigkeit im CCBI-Kontext zwar relevant, da die Netzwerke OLAP-ähnlich arbeiten und damit auch die üblichen Geschwindigkeiten von lokalen Systemen annähernd erreichen sollten,<sup>555</sup> wesentlicher ist aber die Güte des Ergebnisses. Dieser Fokus auf die Datenauswertung zeigt sich auch daran, dass sich in typischen OLAP-Sessions etwa 70% aller Abfragen eine Betrachtungs- und Interpretationszeit von mindestens 45 Sekunden anschließt.<sup>556</sup>

Entscheidend für die Nutzbarkeit von CCBI-Netzwerken ist daher die (messbare) Güte der Ergebnisse bzw. der gelieferten Daten und/oder Informationen, die seit

---

<sup>548</sup> Vgl. Provan/Fish/Sydow (2007): 505f.

<sup>549</sup> Vgl. hierzu auch die Ausführungen zum Thema ‚caching‘ in Abschnitt 4.2.2.

<sup>550</sup> Vgl. Crespo/Garcia-Molina (2005): 2f., 12.

<sup>551</sup> Vgl. die Ausführungen zu Privacy-Überlegungen, Datenhoheit und der allgemeinen Netzstruktur in den Abschnitten 4.4 und 4.5.2.1.

<sup>552</sup> Vgl. Abschnitt 4.1.

<sup>553</sup> Für P2P-Netzwerke liegen – der Natur der Netze entsprechend – keine exakten Zahlen vor. Es lassen sich aber z. B. für das populäre Netzwerk *BitTorrent* mehr als 100 Millionen aktive Nutzer identifizieren (Stand: November 2014, vgl. Torrentfreak (2014)).

<sup>554</sup> Vgl. Barabasi/Albert (1999): 501; Bonifati et al. (2010): 251. Zu den Nutzerzahlen in CCBI-Netzen vgl. Abschnitt 2.4.

<sup>555</sup> Vgl. Abschnitt 3.1.3.

<sup>556</sup> Vgl. Sapia (2000): 226.

mehreren Jahren unabhängig von Netzwerken diskutiert wird. Der folgende Abschnitt bietet deshalb zunächst einen kurzen Exkurs zu dieser Thematik. Im Anschluss werden die relevanten Messgrößen und deren Einflussfaktoren im Bereich von CCBI-Netzwerken diskutiert sowie entsprechende Maßzahlen für die einzelnen Netzwerkbestandteile entwickelt.

### 5.1 Exkurs: Daten- und Informationsqualität

Die *Qualität* (von lat.: *qualitas*) eines Betrachtungsgegenstandes ist dem Ursprung des Wortes nach zunächst die Beschaffenheit oder die Menge der Charakteristika desselben und trifft keine wertende Aussage. Der Begriff ist allerdings im Sprachgebrauch üblicherweise positiv konnotiert und wird auch als ‚Güte‘ verstanden.<sup>557</sup> Stellen Daten den Betrachtungsgegenstand dar, so wird im IT-Kontext von *Datenqualität* (DQ) gesprochen. Das entsprechende Forschungsgebiet erfährt insbesondere seit den 1990er-Jahren vor dem Hintergrund weitentwickelter Datenbanken und Datenaustauschverfahren große Aufmerksamkeit.<sup>558</sup> Als Standardreferenz der Publikationen, die sich mit der Definition des Begriffes auseinandersetzen, kann WANG/STRONG angesehen werden, deren Definition von Datenqualität als „data that are fit for use by data consumers“<sup>559</sup> immer wieder – teilweise leicht abgewandelt – aufgegriffen und verwendet wird.<sup>560</sup> Eine begriffliche Unschärfe stellt in diesem Zusammenhang die üblicherweise bedeutungsidentisch verwendete *Informationsqualität* (IQ) dar, die insbesondere in neueren Veröffentlichungen diskutiert wird, dort dann auch von KAHN/STRONG/WANG.<sup>561</sup> KNIGHT/BURN bezeichnen die Begriffe explizit als vollständig austauschbar.<sup>562</sup> In der vorliegenden Arbeit wurde die Unterscheidung von Daten und Informationen bereits diskutiert.<sup>563</sup> Da bei der Interaktion von Unternehmen auch eine Semantik der ausgetauschten Daten vorliegt und insbesondere im Bereich der Harmonisierung der Daten über das rein Formale hinausgehende Aspekte betrachtet wurden, findet im Weiteren einheitlich der Begriff Informationsqualität Verwendung.

---

<sup>557</sup> Vgl. Duden (2015f).

<sup>558</sup> Vgl. Strong et al. (1994): 500.

<sup>559</sup> Wang/Strong (1996): 6.

<sup>560</sup> Vgl. Knight/Burn (2005): 160; Naumann/Leser/Freytag (1999): 450; Scannapieco/Missier/Batini (2005): 6; Tayi/Ballou (1998): 56.

<sup>561</sup> Vgl. Kahn/Strong/Wang (2002): 184f.; Melkas (2004): 74; Naumann/Leser/Freytag (1999): 450.

<sup>562</sup> Vgl. Knight/Burn (2005): 159.

<sup>563</sup> Vgl. Abschnitt 2.2.4.

Zur Beschreibung und Messung von IQ können mehrere Kriterien oder ‚Dimensionen‘ herangezogen werden, deren Verständnis essenziell für die Beurteilung der Qualität ist.<sup>564</sup> Eine Studie von WANG/STRONG mit datenverarbeitenden IT-Mitarbeitern in Unternehmen verwendet eine initiale Aufteilung in vier Hauptkategorien. Daten müssen demnach aus Empfängersicht

- *korrekt*, d. h. fehlerfrei, objektiv und aus verlässlichen Quellen,
- *zugreifbar*,
- *interpretierbar* und
- *relevant* sein, was eine zeitlich angemessene Verfügbarkeit mit einschließt.<sup>565</sup>

Die Hauptkategorie der Interpretierbarkeit wurde im Anschluss an die Studie in die allgemeine *Darstellung* umbenannt, die Interpretierbarkeit sowie leichte Verständlichkeit mit einschließt.<sup>566,567</sup> Diese vier Kategorien, sowie die entsprechenden Einzelkriterien wurden von diversen Autoren ergänzt und überarbeitet, sodass mehrere Konzepte zur Beschreibung von Datenqualität existieren.<sup>568</sup> SCANNAPIECO/MISSIER/BATINI identifizieren vier Kriterien, die allen betrachteten Konzepten gemein sind:<sup>569</sup>

- *Korrektheit* beschreibt die fehlende Abweichung von einem als ‚richtig‘ angenommenen Wert, wobei sich syntaktische und semantische Korrektheit unterscheiden lassen.
- *Vollständigkeit* beschreibt den Umfang der Daten in Quantität und fachlichem Detail, wobei Referenzmengen existieren können, gegen die geprüft wird.
- *Zeitbezug* beschreibt die Geschwindigkeit der Bereitstellung der Daten, sowohl in Bezug auf den absoluten Zeitraum von Erzeugung zu Auswertbarkeit als auch auf die Häufigkeit der Änderung und die aufgabenbezogene Aktualität der Daten.

---

<sup>564</sup> Vgl. Tayi/Ballou (1998): 56.

<sup>565</sup> Vgl. Wang/Strong (1996): 9.

<sup>566</sup> Vgl. Wang/Strong (1996): 16.

<sup>567</sup> Die Wichtigkeit der Darstellung betont auch ENGLISH, dessen Ausführungen insbesondere den praktischen Einsatz von Datenqualitätsmaßen und -werkzeugen betreffen (vgl. English (2009): 177–186).

<sup>568</sup> Für Aufstellungen und Vergleiche dieser Ansätze vgl. z. B. Knight/Burn (2005): 160–162; Scannapieco/Missier/Batini (2005): 12.

<sup>569</sup> Vgl. zur folgenden Aufzählung Scannapieco/Missier/Batini (2005): 7–12.

- *Konsistenz* beschreibt die semantische Intaktheit der Daten, d. h. die Einhaltung definierter semantischer Regeln, z. B. im Rahmen referentieller Integrität.

Es besteht dabei eine Abhängigkeit der Dimensionen untereinander, die eine Verbesserung der Qualitätswerte einer Dimension ggf. nur unter Inkaufnahme der Verschlechterung von Qualitätswerten einer anderen Dimension ermöglicht. So erfordert z. B. die Bereinigung von Daten im Sinne einer hohen Korrektheit üblicherweise Zeit, was den Zeitbezug der Daten verschlechtert.<sup>570</sup> Die vorgestellten Dimensionen sowie die hier nicht im Detail aufgeführten Ergänzungen lassen sich für den jeweiligen Anwendungskontext passend zusammenstellen, sodass die für die jeweils betrachteten Entitäten relevanten Kriterien eine angemessene Qualität beschreiben.<sup>571</sup> WAND/WANG unterscheiden zudem eine interne und eine externe Sicht auf ein Informationssystem und die jeweilig anzuwendenden IQ-Kriterien. Während interne Kriterien durch das Design des Systems beeinflusst werden, bieten externe Kriterien eine ‚black-box‘-Sicht auf die Leistungsfähigkeit des Systems. Unter diesen sind u. a. auch Informationsgrad, Datenmenge und Detailebene,<sup>572</sup> die intuitiv an der multidimensionalen Sicht eines CCBI-Netzes ausgeprägt werden können.<sup>573</sup>

Eine eigene Betrachtung der bisher nur als Teil der Korrektheit beschriebenen Quellenglaubwürdigkeit der Daten findet sich unter dem Begriff „data provenance“<sup>574</sup>, die z. B. für Datenquellen im WWW besonders relevant ist.<sup>575</sup> In CCBI-Netzen wird dieses Vertrauen in die Daten bereits durch die Überlegungen zum Vertrauen in die Netzteilnehmer behandelt.<sup>576</sup>

---

<sup>570</sup> Vgl. Scannapieco/Missier/Batini (2005): 10.

<sup>571</sup> Vgl. Knight/Burn (2005): 162f., 167f..

<sup>572</sup> Vgl. Wand/Wang (1996): 88, 92.

<sup>573</sup> So bietet der Detailgrad z. B. einen starken Bezug zu der hierarchischen Sicht und der damit verbunden Diskussion um Aggregationen (vgl. Abschnitt 3.1.3 sowie für die weiteren Überlegungen Abschnitt 5.3).

<sup>574</sup> Malaverri/Mota/Medeiros (2013): 2.

<sup>575</sup> Vgl. Prat/Madnick: 1f.

<sup>576</sup> Vgl. Abschnitt 4.5.2.5. Die Einführung einer Metrik zur Glaubwürdigkeit kann dabei möglicherweise die vertrauensbildenden Maßnahmen unterstützen, wird hier aber nicht als eigenes, maßgebliches Kriterium der IQ selbst behandelt.

## 5.2 Mess- und Einflussgrößen in Business-Intelligence-Netzwerken

Während für den allgemeinen Fall eine große Menge an IQ-Messgrößen zur Verfügung steht, ist dies für CCBI-Netzwerke nicht der Fall – insbesondere, da hier nicht ein eindeutiger und monolithischer Datenbestand bewertet wird, sondern die Kombination mehrerer Data Marts und deren Verknüpfung bzw. versendete Auszüge. Im Folgenden wird an einem einfachen CCBI-Netzwerk gezeigt, welche Elemente eines Netzwerks in welcher Art betrachtet werden müssen und welche Messgrößen(arten) dafür grundsätzlich zu erfassen sind. Darauf aufbauend werden allgemeine Netzwerkeigenschaften und -formen erläutert, die Einfluss auf die Messgrößen nehmen. Abschließend wird die Methodik zur Entwicklung dieser erläutert. Es ist bei allen folgenden Überlegungen zu berücksichtigen, dass die zu entwickelnden Messgrößen für ein Netzwerk auch tatsächlich unter den gegebenen Umständen ermittelbar sein sollen, d. h. mit dem – ggf. kumulierten – Wissen der Teilnehmer über das Netzwerk erzeugt werden können und keine darüber hinausgehenden Informationen verlangen.<sup>577</sup>

### 5.2.1 Struktur und Beziehungen der Netzwerkelemente

Abb. 41 zeigt den Graphen eines einfachen Netzwerks, das aus fünf Knoten besteht ( $p_1, p_2, p_3, p_4, p_5$ ). Diese fünf Knoten genügen mit ihren in Summe sechs Verbindungen den Eigenschaften eines CroCoBIN, da sie ein zusammenhängendes Netzwerk bilden und die enthaltenen Zyklen in einem CroCoBIN explizit erlaubt sind.

<sup>577</sup> Der Begriff der Messgröße wurde und wird in dieser Arbeit in seinem allgemeinsprachlichen Verständnis gebraucht als ‚Wert, der die (quantifizierte) Eigenschaft von etwas wiedergibt‘. Die im Weiteren vorgestellten Überlegungen entsprechen diesem Verständnis. Eine dazu grundsätzlich übertragbare formale Definition als Funktion auf einem mathematischen Körper gibt BILLINGSLEY (vgl. Billingsley (1995): 160f.). Vgl. zur Unterscheidung von *Messgrößen* (im Sinne von ‚an einem beliebigen CroCoBIN ohne weitere Kenntnis aus den Informationen der Nutzer ermittelbar‘) und *Kontrollgrößen* (im Sinne von ‚theoretisch bei Vorliegen über das Wissen der Nutzer hinausgehender Daten ermittelbar zur Bewertung der Messgrößen selbst‘) die Ausführungen zur Messgrößenentwicklung in Abschnitt 5.2.3 sowie die entsprechenden Überlegungen in Abschnitt 5.3.



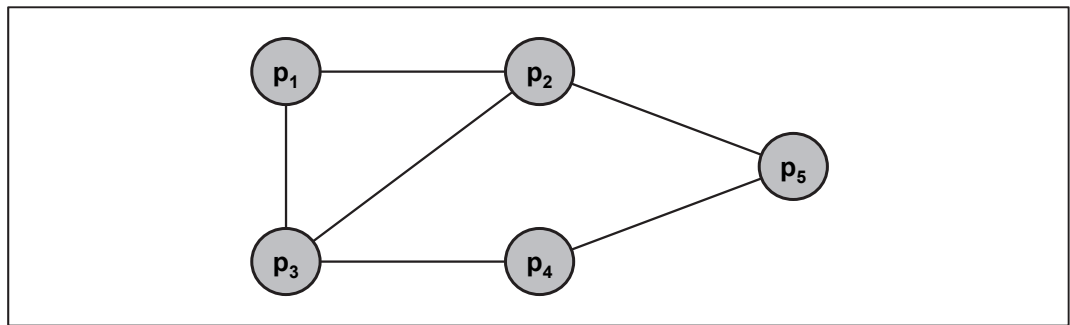


Abb. 41: Graph eines einfachen CroCoBIN

(Quelle: Eigene Darstellung)

Zur einheitlichen Darstellung und Beschreibung dieses und weiterer Netzwerke, wird eine der Graphentheorie entlehnte Notation verwendet, dabei jedoch nur der für das Verständnis dieser Arbeit notwendige Umfang wiedergegeben.<sup>578</sup>

**Definition (13)** Ein *Netzwerk* (Graph)  $G$  besteht aus

- einer Menge an **Peers** (Knoten)  $V = \{v_1, v_2, \dots, v_n\}$  und
- einer Menge an **Verbindungen** (Kanten)  $E = \{e_1, e_2, \dots, e_m\}$ .

**Definition (14)** Jede **Verbindung**  $e$  weist einen Startknoten  $v_i$  und einen Endknoten  $v_j$  auf, mit  $v_i \neq v_j$ , und kann notiert werden als  $e = (v_i, v_j)$  oder vereinfacht als  $v_i v_j$ . Weisen die Kanten  $(v_i, v_j)$  und  $(v_j, v_i)$  unterschiedliche Eigenschaften auf, so ist  $(v_i, v_j)$  eine von  $v_i$  nach  $v_j$  **gerichtete Verbindung**, andernfalls eine **ungerichtete Verbindung**.<sup>579</sup>

<sup>578</sup> Vgl. auch die einleitenden Ausführungen in Abschnitt 4.5.2.1. Für eine umfassende Einführung in die Graphentheorie vgl. z. B. Clark/Holton (1994); Diestel (1996); Gross/Yellen/Zhang (2014). Unter den angegebenen Werken werden für diese Arbeit bei Uneinheitlichkeit die Beiträge in Gross/Yellen/Zhang (2014) als führend angenommen. Die Begriffe Netzwerk und Graph werden, sofern auf die Struktur oder vereinfachte grafische Darstellung des Netzwerks Bezug genommen wird, bedeutungsgleich verwendet. Für die folgenden Definitionen vgl. Gross/Yellen (2014): 2–6, 10.

<sup>579</sup> Ein CroCoBIN ist ohne Berücksichtigung der semantischen Eigenschaften *ungerichtet*, da jede Verbindung zweier Peers immer in beide Richtungen genutzt werden kann. Wird die semantischen Wertigkeit der einzelnen Verbindungen mit einbezogen ist sie jedoch *gerichtet*, da bei abweichenden Schemata in der Regel eine Überführung  $(v_i, v_j)$  nicht vollkommen identisch zu einer Überführung  $(v_j, v_i)$  ist. Für die Abfragen und Antworten wird allerdings eine eindeutige Verbindung genutzt (vgl. Abschnitt 4.5.2.4), sodass nach Festlegung der Kantenbewertung das Netzwerk als ungerichteter Graph betrachtet werden kann. Die Forderung nach der Knotenungleichheit wird hier zur Schaffung eines einfachen Graphen eingebracht.



**Definition (15)** Die zwei Knoten einer Verbindung sind **benachbarte** (adjazente) **Knoten**. Zwei Verbindungen, die mindestens einen Knoten gemeinsam haben, sind **benachbarte Verbindungen**. Die Menge aller benachbarten Knoten von  $v$  ist die Nachbarschaft  $N$  des Knotens.

**Definition (16)** Ein **Pfad** in einem Netzwerk  $G$  ist eine Reihenfolge von Knoten  $v_0, v_1, \dots, v_n$ , wobei für alle Knotenpaare  $v_i, v_{i+1}$  des Pfades gilt, dass  $(v_i, v_{i+1}) \in E$  (mit  $0 \leq i < n$ ), und für alle  $v_i, v_j$  mit  $i \neq j$  des Pfades gilt, dass  $v_i \neq v_j$ .<sup>580</sup> Die Länge eines Pfades ist die Anzahl der Kanten im Pfad. Ein Teilpfad eines Pfades  $P$  ist ein Pfad, dessen Knoten ohne Verlust der Ordnung eine Teilfolge der Knotenfolge von  $P$  darstellen.  $P(v_A, v_B)$  ist der Pfad von  $v_A$  nach  $v_B$ .

Der Vollständigkeit halber zeigt Abb. 42 das exemplarische Netzwerk mit angepasster Notation. Es lassen sich daran vier grundlegende Betrachtungsgegenstände für eine Messgrößenentwicklung erkennen.

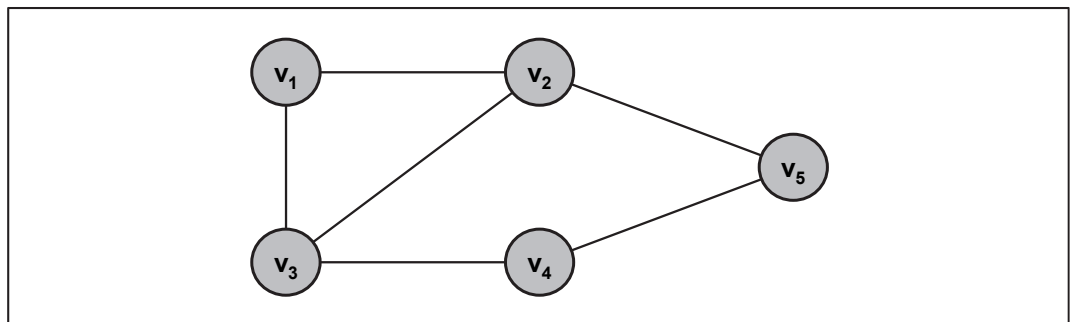


Abb. 42: Graph eines einfachen CroCoBIN in angepasster Notation

(Quelle: Eigene Darstellung)

<sup>580</sup> Diese Definition setzt das Vorliegen eines ‚einfachen‘ Graphen voraus, der keine Mehrfachkanten oder Kanten, bei denen Start- und Endknoten identisch sind (‚Schlinge‘), enthält, was durch Definition (14) und die Erläuterungen zur Definition eines CroCoBIN (vgl. Abschnitt 4.5.2.1) sichergestellt ist. Die Forderung nach der Knotenungleichheit zweier Knoten im Pfad unterscheidet diesen von einem *Weg*, der Zyklen erlaubt. Diese Eigenschaft wird in einem CroCoBIN miteinbezogen, da ein Zyklus in einem Weg bei der Datenübermittlung und -überführung offensichtlich maximal Güte-erhaltend, aber nie -verbessernd sein kann. Der sprachlichen Einfachheit geschuldet werden die Begriffe ‚Pfad‘ und ‚Weg‘ im Weiteren allerdings bedeutungsgleich verwendet.

Die einzelnen *direkten Verbindungen* zwischen den Peers bilden die Basis des Netzwerks. Zu ihrer Beurteilung muss die Qualität des jeweiligen Mappings – im Weiteren: *Mapping-Qualität* (MQ) – herangezogen werden, die bereits im Zusammenhang mit Matching-Verfahren diskutiert wurde.<sup>581</sup> Dabei ist insbesondere zu bestimmen, in welcher Form der Informationserhalt überhaupt definiert werden kann, d. h. welche Eigenschaften der Mappings gemessen werden. In Frage kommen z. B. die Anzahl übereinstimmender Elemente oder die aufrechterhaltene Granularität bei der Überführung. Ein direkter Einfluss auf die MQ ist dabei im Rahmen der CCBI-Betrachtungen nicht vorhanden. Die Datenbestände sind im Vorhinein existent und, weil autonom, auch nicht veränderlich. Für das Matching selbst wird angenommen, dass – unabhängig von dem eingesetzten und im Übrigen beliebigen<sup>582</sup> Verfahren – jede Peer-Paarung das bestmögliche Mapping entwickelt, das auf Basis der vorhandenen Daten und Schemata erzielbar ist. Die MQ ist damit eine objektive, dem Matching nachgelagerte und rein deskriptive Größe.

Ein Großteil der Daten wird zwischen nicht direkt verbundenen Peers über (längere) *Pfade* ausgetauscht. Zur Auswahl des bestmöglichen Pfades müssen alle Möglichkeiten bewertet werden. Im vorgestellten Netzwerk existieren z. B. für die Übertragung von  $v_1$  zu  $v_5$  die alternativen Pfade  $P_1 = v_1, v_2, v_5$  und  $P_2 = v_1, v_3, v_4, v_5$ .<sup>583</sup> Die reine Länge des Pfades trifft dabei allerdings keine Aussage über seine Güte, da die Kombination der Verbindungen im kürzeren Pfad  $P_1$  nicht notwendigerweise mindestens so ‚gut‘ in Bezug auf die Mappings und übertragenen Daten ist wie die Kombination der Verbindungen des Pfades  $P_2$ . Daher ist zur Bestimmung der *Pfad-Qualität* (PQ) eine geeignete Kombination der MQ der Verbindungen im jeweiligen Pfad notwendig.

Im Rahmen des *Routings* werden die zur Datenübertragung tatsächlich zu verwendenden Pfade ausgewählt. Dies geschieht auf Basis einer definierten Routing-Strategie, die festlegt, nach welchen Regeln und Werten die Pfade ausgesucht

<sup>581</sup> Vgl. Abschnitte 4.3 und 4.5.3.

<sup>582</sup> Vgl. die organisatorischen Rahmenbedingungen eines CroCoBIN in Abschnitt 4.5.2.3.

<sup>583</sup> Es existieren noch zwei weitere Pfade, die die Kante  $v_2v_3$ , bzw.  $v_3v_2$  nutzen. Diese sind jedoch irrelevant, da zwar bei mehreren Pfaden der Länge größer oder gleich zwei ohne Anwendung einer konkreten Metrik die Vorteilhaftigkeit nicht direkt erkennbar ist, das Vorliegen einer Metrik (vgl. Definition (12) der Netzwerkstruktur eines CroCoBIN) als solcher aber bereits definiert, dass die direkte Verbindung zweier Knoten  $u$  und  $v$  immer mindestens so gut ist wie ein Pfad von  $u$  nach  $v$  der Länge größer oder gleich zwei. Begründet wird dies durch die *Dreiecksungleichung* als definitorischem Bestandteil einer Metrik (vgl. Walter (2007): 270f.). Vgl. für weitere Ausführungen dazu die Abschnitte 5.3.1–5.3.3.

werden.<sup>584</sup> Die entsprechenden Entscheidungen betreffen dabei das Netzwerk in seiner jeweils aktuellen Konstellation. Die Güte aller verwendeten Pfade wird durch die *Routing-Qualität* (RQ) abgebildet.

Im Fokus der Betrachtung dieser Arbeit steht abschließend das *Netzwerk als Ganzes*, für das beurteilt werden soll, in welchem Maße es geeignet ist, jedem Teilnehmer umfassende und detaillierte Informationen auf Basis aller verfügbaren Teilnehmerdaten bereitzustellen. Dabei steigt – unter der Annahme, dass ‚gute‘ Routings eine hohe Informationsgüte hervorbringen – die *Netzqualität* (NQ) in der Regel mit der Güte aller im Netzwerk verwendeten Routings. Das Netzwerk ist die Summe der Teilnehmer und ihrer Verbindungen. Es verfügt im Ganzen über keine parametrisierbaren Größen, da für CroCoBIN ausschließlich die fachlichen Aspekte betrachtet werden und keine Restriktion in Bezug auf z. B. technische Ressourcen (physische Verbindungen, etc.) angenommen werden. Es existieren wohl aber indirekte Einflussfaktoren, die aus netzspezifischen Rahmenbedingungen resultieren können. So kann z. B. eine Zielverbindungsanzahl je Knoten definiert werden.<sup>585</sup> Umfangreichere Bedingungen lassen sich durch die Definition einer (Ziel-)Topologie, d. h. einer übergreifenden Netzstruktur, schaffen, die auch in nicht-technischen Netzen zur Entscheidungsfindung erheblichen Einfluss auf die Zufriedenheit der Teilnehmer und die Qualität des Ergebnisses hat.<sup>586</sup> Liegen derartige indirekte Einflussfaktoren vor, können Verbindungen ggf. nicht vollständig frei gewählt werden, d. h. die *Nachbarschaftswahl* – aus Sicht eines neu in das Netzwerk eintretenden Peers – ist eingeschränkt.<sup>587</sup> Diese beeinflusst aber maßgeblich die Routing-Qualitäten, da ggf. nicht alle Verbindungen zur Verfügung stehen, die bei Auswahl ‚anderer‘ Nachbarn zur Verfügung stünden. Maßgeblicher Einflussfaktor auf die Nachbarschaftswahl ist – neben den Rahmenbedingungen – wiederum die MQ, mit der die (neuen) Netzteilnehmer vielverspre-

---

<sup>584</sup> Vgl. Abschnitt 4.5.2.4.

<sup>585</sup> Vgl. Abschnitt 4.5.2.5.

<sup>586</sup> Vgl. Swap (1984): 56f. Zu verschiedenen Topologien und weiteren, allgemeinen Netzeigenschaften vgl. Abschnitt 5.2.2.

<sup>587</sup> Die Nachbarschaftswahl erfolgt immer, wenn ein neuer Knoten dem Netzwerk beitrifft. Verlässt hingegen ein Knoten das Netz, so ist dies entweder ‚folgenlos‘ für die bestehenden Routings – z. B. weil der entfernte Knoten in keinem Routing Abfragen weitergeben musste – oder bestehende Knoten müssen diesen Knoten in ihren Routings ‚ersetzen‘, was entweder durch die Auswahl anderer Pfade oder durch die Um-/Neubildung von Nachbarschaften geschieht, sodass die Problematik grundsätzlich identisch bleibt. Im Sinne der Überlegungen zur RQ können die Fälle daher einheitlich behandelt werden.

chende Partner ausmachen können.<sup>588</sup> Abb. 43 fasst die Zusammenhänge der Elemente, Kriterien und Einflussfaktoren zusammen, auf deren Basis die tatsächlichen Messgrößen entwickelt werden.


Element / Ebene	Messgröße			Mögliche Einflussfaktoren auf die Güte des Netzes
	Name	Gegenstand	Einflussgrößen	
<b>Netz</b>	<b>NQ</b>	Güte aller in einem Netz verwendeten Routings	Kombination der Routing-Qualitäten Weitere Faktoren (z. B. Robustheit)	Rahmenbedingungen (z. B. Mindestanzahl an Verbindungen, Ziel-Topologie)
<b>Routing</b>	<b>RQ</b>	Güte aller von einem Knoten tatsächlich verwendeten Pfade	Routing-Strategie (Beurteilung und Auswahl der verfügbaren Pfade)	Nachbarschaftswahl
<b>Pfad (Weg)</b>	<b>PQ</b>	Güte der Kombination von Verbindungen in einem Pfad	Kombinationsfunktion für die Mapping-Qualitäten Verwendung gerichteter oder ungerichteter Kanten für die Mapping-Qualitäten	
<b>Mapping (Verbindung)</b>	<b>MQ</b>	Güte eines Mappings zweier Peers	Kriterien für den Aufbau der Mapping-Qualität (z.B. Anzahl identischer Elemente, Granularität)	Prognostizierte / Erwartete Mapping-Qualität
Legende  Nimmt Einfluss auf				

Abb. 43: Schematischer Zusammenhang von Netzwerkelementen, Messgrößen und Einflussfaktoren zur Qualitätsmessung in einem CroCoBIN  
(Quelle: Eigene Darstellung)

## 5.2.2 Aufbau und Eigenschaften von Netzwerken

Bei dem Aufbau von IT-gestützten Netzwerken werden typischerweise mehrere idealisierte Netzwerk-Topologien unterschieden, die die Grundstruktur der ver-

<sup>588</sup> Kritisch zu betrachten ist die hier implizierte Kenntnis von der Güte eines noch nicht geschaffenen Mappings. In der praktischen Umsetzung kann üblicherweise – aus Gründen zu hohen Aufwands, alle möglichen Kombinationen abzubilden – lediglich eine Prognose der zu erwartenden MQ zu unterschiedlichen Partnern abgegeben werden, z. B. auf Basis von automatisierten Matching-Verfahren. Für die Betrachtung der Messgrößen und die im Weiteren vorgenommenen Überlegungen zur Netzgestaltung (vgl. Kapitel 6) ist die Annahme der MQ-Kennntnis allerdings plausibel, da näherungsweise verwendete Werte aus automatisierten Verfahren nur die Güte des Ergebnisses von einzelnen Implementierungen beeinflussen und nicht die der idealtypischen Vorgehensweise. Im Übrigen kann die Güte der MQ-Prognose selbst als Variable dienen, sodass betrachtet werden kann, welche Auswirkungen auf die NQ unterschiedlich hohes Wissen über das Netzwerk vor dem Eintritt eines Knotens in das Netzwerk hat (vgl. Abschnitte 5.4, 6.4 und 7.2). Unterstützend gelten können für die ursprüngliche Annahme zudem vergleichbare Diskussionen zur Anwendbarkeit in der Realität (äußerst) selten anzutreffender Annahmen und Konzepte zur Modellierung komplexer Fragestellungen mit auch in nicht-idealer Umgebung anwendbaren Ergebnissen, wie sie sich z. B. in der Volkswirtschaftslehre mit dem Konzept eines ‚vollkommenen Marktes‘ seit langer Zeit finden (vgl. Makowski/Ostroy (2001): 479–485; Stigler (1957): 14–17).

bundenen Systeme beschreiben. Abb. 44 stellt die – nach MEINEL/SACK – vier etablierten Formen dar. Diese können auch kombiniert oder erweitert werden.

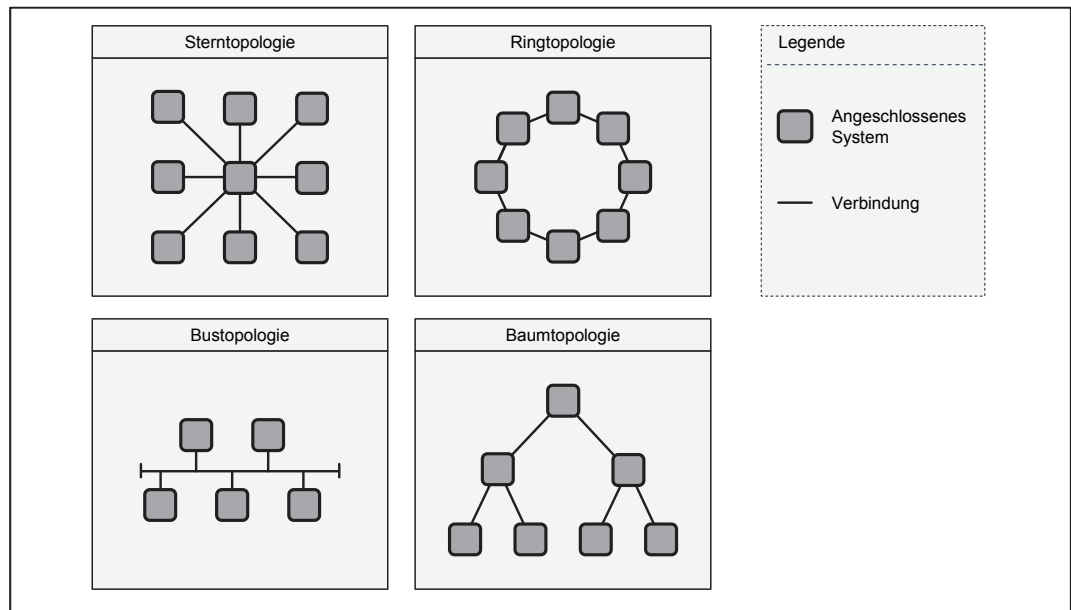


Abb. 44: Vergleich unterschiedlicher Netzwerk-Topologien

(Quelle: In Anlehnung an Meinel/Sack (2012): 174)

Die Festlegung einer Netzwerktopologie wird in einem CroCoBIN allerdings nicht über technische Verknüpfungen durchgeführt, sondern durch die Auswahl der Partner zur Überführung von Daten. Sofern mehrere Partner gewählt werden, existiert damit eine – zumindest etwas – komplexere Topologie, die die möglichen Verbindungen beschreibt. Die *tatsächliche Topologie* in der Nutzung kann von dieser *ursprünglichen Topologie* jedoch abweichen, da bei alternativen Routing-Möglichkeiten ggf. einzelne Verbindungen aufgrund ihrer geringen Güte immer ignoriert werden. Alle Topologien weisen unterschiedliche Eigenschaften auf, die nach mehreren Kriterien beurteilt werden können, sodass eine dem Einsatzzweck angemessene Auswahl stattfinden kann. Neben der *Effizienz* (geringer Aufwand für den Datentransport) und *Robustheit* (Bewahrung der Funktionsfähigkeit bei Ausfall einzelner oder mehrerer Knoten) existieren für IT-Netzwerke noch der *Verkabelungsaufwand*, der in einem CroCoBIN mit dem Matchingaufwand zur Erstellung eines Mappings zu allen Partnern gleichgesetzt werden kann und die *Gesamtbandbreite*, die für Teilstrecken den maximalen Durchsatz definiert, was

in einem CroCoBIN durch die Aufrechterhaltung der Abfrage- und Antwortqualität abgebildet wird.<sup>589</sup>

Die Kriterien für die Auswahl der Topologie korrespondieren in Teilen mit den Anforderungen an P2P-Netze, zu denen ebenfalls die Effizienz und Robustheit zählen, die durch *Skalierbarkeit* (das problemlose Hinzufügen weiterer Knoten bei Aufrechterhaltung der grundlegenden Eigenschaften) und *Selbstorganisationsfähigkeit* (Abwesenheit einer zentralen Kontrollinstanz) ergänzt werden.<sup>590</sup> Um die Frage nach der Ausprägung der Kriterien und Erfüllung der Anforderungen in einer nicht-trivialen Topologie beantworten zu können, werden unterschiedliche Eigenschaften von Netzwerken betrachtet. Dazu gehören bspw. die (durchschnittliche) Anzahl der Verbindungen je Knoten oder die Länge maximaler Pfade im Netzwerk. Maßgeblich für die Betrachtung einzelner Eigenschaften ist jedoch ihr Nutzen für die Ableitung von Aussagen über das Netzwerk.<sup>591</sup> So lassen sich – insbesondere für Kommunikationsnetzwerke – Aussagen über die Robustheit gegenüber zufälligem Ausfall und bewusster Sabotage treffen.<sup>592</sup> Als grundlegende Netzwerkeigenschaften, bzw. Grapheneigenschaften, lassen sich dabei die folgenden ausmachen.<sup>593</sup>

**Definition (17)** Der *Grad* eines Knotens  $d(v)$  ist die Anzahl der Nachbarn und damit die Größe der Nachbarschaft  $N$ .

**Definition (18)** Die *Dichte* eines Graphen  $d(G)$  ist das Verhältnis der Anzahl existierenden Verbindungen zu der Anzahl möglicher Verbindungen:  $d(G) = \frac{2|E|}{|V|(|V|-1)}$ .

<sup>589</sup> Vgl. zu den vorgestellten Eigenschaften der IT-Netze Meinel/Sack (2012): 174f.

<sup>590</sup> Vgl. Aberer et al. (2005): 12.

<sup>591</sup> Vgl. Provan/Fish/Sydow (2007): 484f. So können z. B. für soziale Netzwerke diverse Eigenschaften ausgemacht werden, u. a. die stark heterogene Anzahl von Verbindungen der Teilnehmer, aus denen Rückschlüsse über die Nutzer und ihr Verhalten gezogen werden können (vgl. Krivitsky et al. (2009): 204).

<sup>592</sup> Vgl. Albert/Jeong/Barabasi (2000): 378–381.

<sup>593</sup> Vgl. zu den folgenden Definitionen Gross/Yellen (2014): 8, 11, 17, 18; Jackson (2008): 29–34; 37–43.

**Definition (19)** Die *Distanz* zweier Knoten  $d(u, v)$  ist die Länge des kürzesten Pfades zwischen ihnen.<sup>594</sup> Die *durchschnittliche Distanz* des Graphen ist das arithmetische Mittel aller Distanzen.

**Definition (20)** Die *Exzentrizität* eines Knotens ist seine maximale Distanz zu allen anderen Knoten des Graphen. Der *Radius* eines Graphen entspricht dann der minimalen, der *Umfang* der maximalen Exzentrizität in ihm.<sup>595</sup>

Ergänzend zu den gegebenen erfolgt die folgende, von der Norm abweichende Definition:

**Definition (21)** Die *Größe* eines Netzwerks ist die Anzahl seiner Teilnehmer.<sup>596</sup>

Die Auswertung der vorgestellten Größen für existierende Netzwerke hat gezeigt, dass, entgegen früherer Annahmen, eine Vielzahl an Netzwerken keine gleichmäßige Verteilung der Verbindungen aufweist.<sup>597</sup> Zur Ermittlung der Verbindungsverteilung wird die Verteilungsfunktion des Knotengrades  $P(k)$  herangezogen, die für ein betrachtetes Netzwerk die Wahrscheinlichkeit beschreibt, bei einem beliebigen Knoten  $k$  Nachbarn vorzufinden. In *Zufallsnetzwerken* (z. B. dem Fernstraßennetzwerk der USA) weist ein Großteil der Knoten das gleiche  $k$  oder eine sehr ähnliche Anzahl auf, während z. B. Flugliniennetzwerke üblicherweise wenige, sehr stark vernetzte ‚Hubs‘ aufweisen. In diesen *skalenfreien Netzwerken* liegt der erwartete Grad für andere Knoten sehr deutlich darunter.<sup>598</sup> Abb. 45 zeigt ein exemplarisches Zufallsnetzwerk im Vergleich zu einem skalenfreien Netzwerk. Die Hubs sind dort grau eingefärbt.

<sup>594</sup> Grad, Dichte und Distanz unterscheiden sich in der Notation immer und notwendigerweise in der Anzahl und Art der Funktionsparameter von  $d$  und sind somit unterscheidbar.

<sup>595</sup> Bei der Betrachtung von Distanz, Exzentrizität und den daraus abgeleiteten Größen ist zu beachten, dass hier nur die existierenden Verbindungen ohne ihre Wertigkeit betrachtet werden, d. h. die ursprüngliche Topologie des Netzwerks beschrieben wird. Um die Maße in einem CroCoBIN auf die tatsächliche Topologie zu beziehen, müssen die Maße nicht auf dem kürzesten, sondern auf dem ‚besten‘, d. h. dem tatsächlich genutzten Pfad basieren.

<sup>596</sup> In der Graphentheorie gilt als Größe üblicherweise die Anzahl der Verbindungen, während die Anzahl der Knoten als *Ordnung* bezeichnet wird (vgl. Gould (2014): 315). Dem allgemeinen Sprachgebrauch folgend – ein Netzwerk ‚wächst‘ mit der Anzahl der Teilnehmer, also mit der ‚Größe der Gruppe der Entscheider‘ – wird hier jedoch die intuitivere Gleichsetzung von Teilnehmerzahl und Größe gewählt.

<sup>597</sup> Vgl. Barabási/Bonabeau (2003): 52.

<sup>598</sup> Vgl. Barabási/Bonabeau (2003): 53.



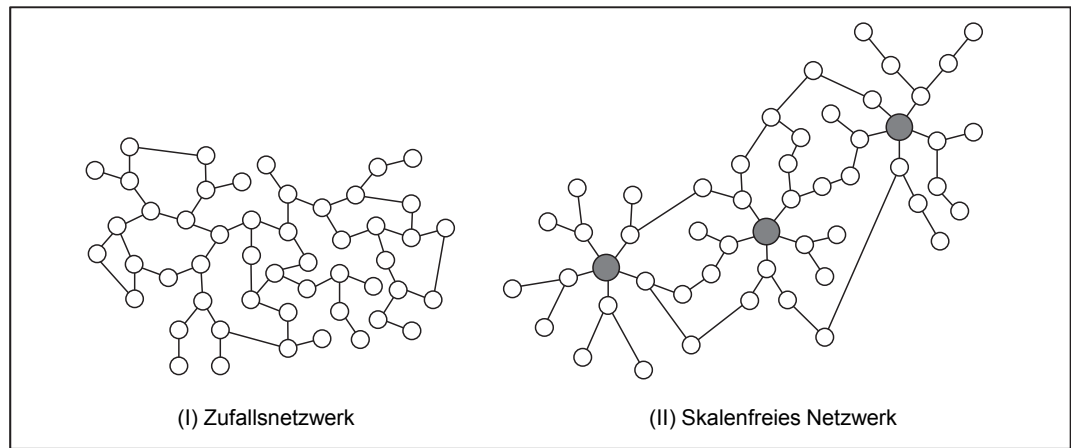


Abb. 45: Vergleich eines Zufallsnetzwerks mit einem Skalenfreien Netzwerk  
(Quelle: Seo et al. (2013): 1739)

Die Wahrscheinlichkeiten für das Vorliegen eines Knotens mit dem Grad  $k$  können in Zufallsnetzwerken durch eine Poisson-Verteilung der Form

$$P_{\lambda}(k) \sim \frac{\lambda^k}{k!} e^{-\lambda}$$

beschrieben werden, wobei  $\lambda$  einen zu ermittelnden Wert besitzt. Skalenfreie Netzwerke hingegen weisen eine Potenzgesetz-basierte Verteilung auf, wobei in

$$P(k) \sim k^{-\gamma}$$

der Wert für  $\gamma$  zu ermitteln ist.<sup>599</sup> Abb. 46 zeigt die Verteilungen im Vergleich.

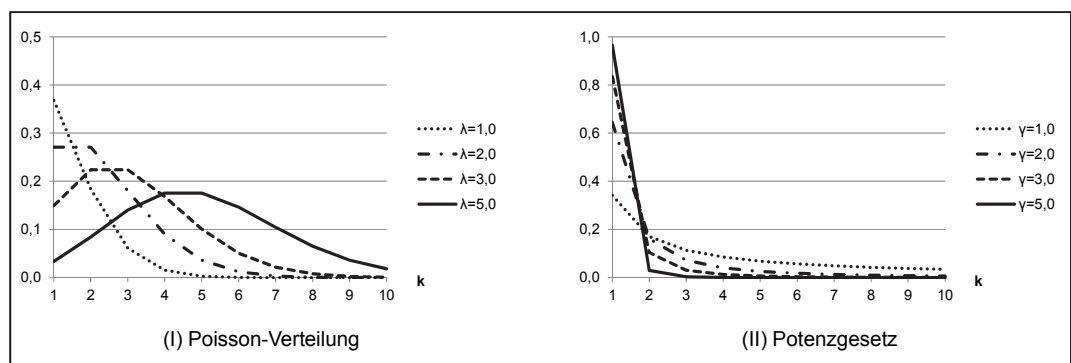


Abb. 46: Werte von Poisson-Verteilungen und Potenzgesetzen abhängig von  $k$ <sup>600</sup>  
(Quelle: Eigene Darstellung)

<sup>599</sup> Vgl. Seo et al. (2013): 1739. Studien weisen für die meisten Netzwerke auf einen Wert  $2 \leq \gamma \leq 3$  hin (vgl. Barabási/Bonabeau (2003): 54).

<sup>600</sup> Die Potenzgesetze bilden keine eigenständige Verteilung, vielmehr folgt die Verteilung einem – näher zu bestimmenden – Potenzgesetz. Für die Darstellung wurden die exemplarischen Wertedarstellungen normiert. Die Poisson-Verteilung weist einen Wert für 0 auf, der hier nicht dargestellt wird, da  $k = 0$  Nachbarn in CroCoBIN-Netzwerken per Definition nicht vorkommen. Die Werte für  $\lambda$  und  $\gamma$  sind in Anlehnung an die typischen Werte gewählt.



Skalenfreie Netzwerke weisen dabei in Bezug auf die Robustheit deutlich bessere Eigenschaften auf als Zufallsnetzwerke. Der Umfang eines Netzwerks erhöht sich bei der Entfernung einzelner Knoten nicht oder kaum, sofern die Knoten zufällig entfernt werden.<sup>601</sup> Durch die Wichtigkeit der Hubs im Netzwerk stellen diese jedoch kritische Angriffspunkte dar, was skalenfreie Netze anfälliger gegenüber bewussten Knotenentfernungen macht.<sup>602</sup> Für CCBI-Netzwerke lässt sich daher prüfen, ob die Forcierung eines skalenfreien Netzes – oder anderer Netzwerkeigenschaften – vorteilhaft erscheint, welche ‚kritische Größe‘ dafür erreicht sein müsste und welche Umweltbedingungen, im Sinne von Manipulationswahrscheinlichkeiten, vorliegen müssen. Es zeigt sich in jedem Fall, dass die Betrachtung der allgemeinen Netzwerkeigenschaften Teil und Vorstufe der Entwicklung von netzspezifischen Messgrößen ist.<sup>603</sup>

### 5.2.3 Methodik zur Erarbeitung von Messgrößen

Da CCBI-Netzwerke bisher nicht umfassend in der Literatur diskutiert wurden, existieren keine verwendbaren oder als Vorlage nutzbaren Messgrößen, die über reine Überlegungen zur Qualitätsmessung von Mappings hinausgehen. Im Bereich der PDMS lassen sich jedoch ähnliche Grundannahmen und Fragestellungen finden.<sup>604</sup> Bei einer Abfrage von einem Peer an einen anderen entsteht – sofern kein perfektes Mapping vorliegt – ein Informationsverlust, der als „semantic loss“<sup>605</sup> bezeichnet wird. Da eine Abfrage in P2P-Netzen üblicherweise über einen Pfad mit mehr als zwei Knoten verschickt wird, kann sich dieser Verlust über mehrere Transfers vergrößern. Dieser Effekt wird als Informationsabbau oder „semantic degradation“<sup>606</sup> bezeichnet.<sup>607</sup> Zur Messung eines Informationsverlusts kann zu-

---

<sup>601</sup> Vgl. Albert/Jeong/Barabasi (2000): 379. Diese Annahme kann bei großen Netzwerken ( $|V| \geq 1000$ ) empirisch nachgewiesen werden. Je kleiner das Netzwerk ist, desto höher ist im Einzelfall die ‚Chance‘ einen Hub zu entfernen und den Umfang zu vergrößern.

<sup>602</sup> Vgl. Albert/Jeong/Barabasi (2000): 381.

<sup>603</sup> Nicht betrachtet wurde hier die *Zentralität* als typische Kennzahl in der Netzwerkanalyse. Diese stellt bei der Identifikation von einflussreichen Netzteilnehmern eine maßgebliche Größe dar und wird insbesondere in der Forschung zu sozialen Netzwerken verwendet. Sie ist allerdings individualfokussiert und dient nicht vornehmlich der Erfassung der Netzstruktur im Ganzen (vgl. Ahlf (2013): 25f.). Sie ist zudem in diversen verschiedenen Ausprägungen definiert und kann nur kontextabhängig betrachtet werden. Vgl. für eine Darstellung der unterschiedlichen Varianten und der damit verbundenen Interpretationsschwierigkeiten Borgatti/Everett (2006): 466–484; Knoke/Yang (2008): 62–72.

<sup>604</sup> Vgl. Abschnitt 4.3.4.

<sup>605</sup> Souza/Batista/Salgado (2012): 220.

<sup>606</sup> Lodi et al. (2008): 26.

nächst auf einfache Größen zurückgegriffen werden, die im *Information Retrieval* (IR)<sup>608</sup> zur Bestimmung der IQ bei Vorliegen irrelevanter Daten eingesetzt werden.<sup>609</sup> Diese werden bereits seit mehreren Jahrzehnten eingesetzt und basieren auf dem Vergleich von erhaltenen zu nicht erhaltenen bzw. erwünschten zu nicht erwünschten Daten.<sup>610</sup> Dabei haben sich einzelne Messgrößen wenig aussagekräftig gezeigt, sodass die Kombination in – teilweise sehr komplexen – Konstrukten sinnvoll und üblich ist.<sup>611</sup> Diese Überlegungen finden sich in einer großen Zahl aktueller Ansätze zur Messung einzelner Qualitäten in zu CCBI-Netzen verwandten Informationssystemstrukturen wieder. Messgrößen werden dabei sowohl zur Bestimmung einzelner Bestandteile als auch zur aggregierten Betrachtung kombiniert, sodass eine Messgrößenhierarchie entsteht, wie sie auch in einem CroCoBIN durch die Unterteilung und Bezugnahme der Messgrößen auf mehreren Ebenen beschrieben wurde.<sup>612</sup> Dabei sind diese Größen in Teilen rein additiv, zumeist aber als Durchschnitt gebildet, welcher üblicherweise durch einen gegebenen oder vom Anwender zu bestimmenden Faktorenvektor gewichtet wird.<sup>613</sup>

Die angegebenen Publikationen verwenden zur Ermittlung der Nützlichkeit der entwickelten Messgrößen in der Regel systemische Implementierungen, die prototypisch den Anwendungskontext abbilden und z. B. ein PDMS oder Funktionen in diesem simulieren,<sup>614</sup> was auch dem Vorgehen der vorliegenden Arbeit entspricht.<sup>615</sup> An den bestehenden Arbeiten zeigen sich allerdings auch problematische Aspekte des Vorgehens. So setzen LODI ET AL. eine selbstentwickelte Messgröße zur Bestimmung guter Nachbarschaften ein und zeigen, dass diese zu einer höheren Qualität der Abfragen führt, geben jedoch nicht an, wie sie gemessen

---

<sup>607</sup> Die Bezeichnungen variieren dabei leicht. So existiert alternativ der Begriff „query degradation“ (Souza/Batista/Salgado (2012): 220). „Semantic loss“ wird an anderer Stelle nicht als der Verlust einer Übertragung, sondern einer Hin- und Rückübertragung definiert (vgl. Delveroudis/Lekeas (2007): 63).

<sup>608</sup> Ziel des IR ist die Informationsgewinnung auf Basis einer Abfrage auf einen gegebenen Bestand bei gleichzeitiger Vermeidung einer Rückgabe für die Abfrage irrelevanter Daten (vgl. Rijsbergen (1979): 3f.).

<sup>609</sup> Vgl. Mena et al. (2000): 412.

<sup>610</sup> Vgl. Rijsbergen (1979): 114f. Auch der in Abschnitt 4.3.3 vorgestellte Jaccard-Koeffizient lässt sich auf diese Überlegungen zurückführen.

<sup>611</sup> Vgl. Rijsbergen (1979): 119.

<sup>612</sup> Vgl. Abschnitt 5.2.1.

<sup>613</sup> Vgl. Aligon et al. (2014b): 474; Aouiche/Jouve/Darmont (2006): 86f.; Banek et al. (2007): 51; Duchateau/Bellahsene (2010): 266; Freire et al. (2012).

<sup>614</sup> Vgl. Lodi et al. (2008): 32–35; Mandreoli et al. (2006): 23–25.

<sup>615</sup> Vgl. Abschnitt 1.2.2.

wird.<sup>616</sup> MANDREOLI ET AL. bestimmen mit Hilfe eines selbstentwickelten Index ein gutes Routing in einem PDMS und messen die Güte des Algorithmus, der ein diesen Index optimierendes Netzwerk erzeugt, dann wiederum an den Indizes der verwendeten Verbindungen.<sup>617</sup> Eine objektive Betrachtung ist damit nicht möglich, da der Index sowohl abhängige als auch unabhängige Variable ist, was einer Grundvoraussetzung für den Einsatz von Verfahren zur Erklärung und Prognose von Zusammenhängen widerspricht.<sup>618</sup> Zur Vermeidung dieser ‚selbsterfüllenden Prophezeiungen‘<sup>619</sup> wird daher in der vorliegenden Arbeit getrennt zwischen der Entwicklung der eigentlichen Messgrößen (zu denen auch die NQ gehört) sowie den zu untersuchenden Einflussfaktoren auf die Messgrößen. Der Bestimmung sinnvoller Netzwerk-Konfigurationen gehen in der prototypischen Implementierung auf Kontrollgrößen basierende Vorstudien zur konkreten Ausgestaltung (z. B. der Faktorengewichtung) der Messgrößen voraus, sodass diese im Anschluss als abhängige Variablen für die Effektbestimmung der verschiedenen Konfigurationen verwendet werden können.<sup>620</sup>

Weiterhin ist zu berücksichtigen, dass bei zusammengesetzten Messgrößen – auch nach transparenter Bestimmung von Gewichtungsfaktoren – ggf. nicht bekannt ist, welche statistischen Werte der Verteilungsfunktion der Messgröße vorliegen oder welcher Wert in der Interpretation der Größe als ‚gut‘ oder ‚schlecht‘ angesehen werden sollte. Solange die Größe jedoch gleich definiert bleibt und in vergleichbaren Umgebungen eingesetzt wird – wie es hier in CroCoBIN-Netzwerken der Fall ist – ist sie für einen Vergleich verschiedener Konstellationen der unabhäangi-

---

<sup>616</sup> Die Autoren definieren ‚satisfaction‘ als Wert proportional zur Güte der Abfragen und nutzen diesen mit unbekannter Skala und Herleitung zum Vergleich ihrer Methode und anderer Ansätze (vgl. Lodi et al. (2008): 34).

<sup>617</sup> Die Autoren verwenden auch hier eine ‚satisfaction‘, erläutern aber noch, dass jeder Beitrag zum Wert dieser auf Basis der Mapping-Qualitäts-Werte, d. h. auf Basis des Index, bestimmt wird (vgl. Mandreoli et al. (2006): 24.).

<sup>618</sup> Vgl. Backhaus (2011): 14.

<sup>619</sup> Die ‚selbsterfüllende Prophezeiung‘ entstammt im Ursprung der Philosophie: „Eine sich selbst erfüllende Prophezeiung ist eine Annahme oder Voraussage, die rein aus der Tatsache heraus, daß sie gemacht wurde, das angenommene, erwartete oder vorhergesagte Ereignis zur Wirklichkeit werden läßt und so ihre eigene »Richtigkeit« bestätigt.“ (vgl. Watzlawick (1981): 91). Übertragen auf das gegebene Beispiel wird ein auf die Messgröße optimiertes Netzwerk allein schon deshalb eine gute Qualität ausweisen, weil die Qualität eben genau durch die Höhe der Messgröße bestimmt wird.

<sup>620</sup> Vgl. Abschnitte 5.3 und 5.4 sowie die Ausführungen zur Simulation in Kapitel 6. Eine Kontrollgröße kann z. B. bei vollständiger Kenntnis über Schemata und Datenbestand aller Teilnehmer für die RQ bestimmen, welche RQ maximal erzielbar wäre, wenn nicht einzelne Knoten die Abfragen auf Basis ihrer lokalen Kenntnis weiterleiten, sondern ein global ‚bester‘ Pfad verwendet würde (vgl. dazu die Ausführungen in Abschnitt 5.3.3).

gen Variablen als Vergleichsgröße einsetzbar.<sup>621</sup> Die im Folgenden entwickelten Messgrößen sind daher für die Erreichung des Erkenntnisziels dieser Arbeit anwendbar.<sup>622</sup>

### 5.3 Definition der Messgrößen

Die folgenden Abschnitte beschreiben die Entwicklung der Messgrößen (und ihrer jeweiligen Kontrollgrößen) auf den verschiedenen identifizierten Ebenen des Netzes. Dabei werden die Herleitungen dargestellt und allgemeine Aussagen getroffen. Die konkrete Überprüfung unterschiedlicher Varianten oder die Ausgestaltung tatsächlicher Gewichtungen wird im Rahmen der prototypischen Umsetzung diskutiert.<sup>623</sup>

#### 5.3.1 Mapping-Qualität

Die Ähnlichkeitsbestimmung von Schemata wurde bereits im Rahmen der Harmonisierung multidimensionaler Datenstrukturen betrachtet.<sup>624</sup> Bisherige Ansätze beziehen sich dabei häufig auf die Integration mehrerer Schemata und verwenden Maße zur Bestimmung der gleichartigen Abdeckung eines integrierten Zieldatenbestandes durch unterschiedliche Schemata, z. B. deren Komplettheit oder Konsistenz.<sup>625</sup> Die bisherigen Ansätze prüfen jedoch fast immer ein automatisch generiertes Mapping gegen ein vom Menschen erstelltes und per Definition als ‚perfekt‘ angenommenes Mapping oder ‚Master-Schema‘.<sup>626</sup> Sofern ein solches existiert, erübrigt sich die Frage der Betrachtungsrichtung, da nur das generierte Schema geprüft und bewertet wird. Im Kontext von P2P-Systemen ist die Richtung jedoch ein entscheidendes Kriterium, da der Informationsfluss bei der Bewertung gerichteter Verbindungen Empfänger- oder Sender-orientiert beeinflusst werden kann.<sup>627</sup> Die in einem CroCoBIN verwendete MQ muss daher, trotz der

---

<sup>621</sup> Vgl. Rijsbergen (1979): 135f.

<sup>622</sup> Vgl. Abschnitt 1.1. Die Aussage bezieht sich hier insbesondere auf den zweiten Teil des Erkenntnisziels, ‚gute‘ Werte der Einflussfaktoren und geeignete Rahmenbedingungen für den erfolgreichen Einsatz für CCBI-Netzwerke zu identifizieren.

<sup>623</sup> Vgl. Kapitel 6.

<sup>624</sup> Vgl. Abschnitt 4.3.

<sup>625</sup> Vgl. Batista/Salgado (2007): 62.

<sup>626</sup> Vgl. Duchateau/Bellahsene (2010): 261f.; Yatskevich et al. (2007): 3f.

<sup>627</sup> Die Problematik (un-)gerichteter Verbindungen wird im Rahmen der PQ- und RQ-Betrachtungen vertieft (vgl. Abschnitte 5.3.2 und 5.3.3). Es sei hier vorweggenommen, dass die in anderen Publikationen vorausgesetzte Kommutativität der MQ-Funktion – d. h.,  $MQ(v_i, v_j) = MQ(v_j, v_i)$  (vgl. Mandreoli et al. (2006): 21) – erhebliche Einschränkungen der Bestimmung der zu nutzenden Verbindungen verursacht.

Forderung nach einer eindeutigen Verbindungsreihenfolge für Hin- und Rückweg zwischen zwei Peers,<sup>628</sup> fähig sein, die Richtung der Datenübertragung zu berücksichtigen.<sup>629</sup> In diesem Fall tritt der jeweils abfragende – und Antwortempfangende – Peer als ‚Master‘ auf, dessen Datenbestand auf Basis seines Schemas möglichst gut bedient werden sollte. Ein Mapping ist aus seiner Sicht besonders gut, wenn möglichst viele der theoretisch verfügbaren Zellen des Datenwürfels – aufgespannt durch die im Schema definierten Dimensionen und Fakten – gefüllt werden können, während eine Überdeckung zunächst aus Sicht des Masters irrelevant ist. Dies entspricht im Wesentlichen der Definition des *Recall* aus dem IR.<sup>630</sup> Überprüfbar ist dieser jedoch nur nach einem vollständig durchgeführten Matching. Durch den Bezug auf einzelne Datenpunkte lässt sich zudem nur für jeden Einzelfall eine Aussage über die Qualität der Überführung treffen, da die Schema-Eigenschaften wie z. B. hierarchische Strukturen nicht berücksichtigt werden und die Güte nur vor dem Hintergrund der Peer-Semantik beurteilbar ist. Für jede Verbindung stellt diese *tatsächliche Mapping-Qualität* (TMQ) somit die individuelle Kontrollgröße der MQ dar.

Zur Erstellung einer allgemeineren, schema-basierten MQ können zunächst die bereits im Kontext DWH-spezifischer Harmonisierungsmethoden vorgestellten Eigenschaften eines Dimensions-Mappings verwendet werden.<sup>631</sup>

Die Soundness eines Mappings beschreibt den auf Elementebene vorhandenen Informationserhalt bei der Übertragung. Die von TORLONE als ‚vorhanden/nicht vorhanden‘ beschriebene Größe kann dabei auch anteilig errechnet werden, indem die Anzahl der ‚sound members‘, also der eindeutig übertragbaren Elemente, in Verhältnis zu allen potenziell übertragbaren Elementen (aus Empfängersicht) gesetzt wird:

---

<sup>628</sup> Vgl. Abschnitt 4.5.2.4.

<sup>629</sup> Die MQ implementiert die im CroCoBIN-Referenzmodell definierte Metrik *s* zur Ähnlichkeitsbestimmung (vgl. Definition (12) und Abschnitt 4.5.2.1). Zur einheitlichen Verwendung der Begriffe wird im Folgenden ausschließlich die Bezeichnung MQ verwendet, wenn die Metrik des CroCoBIN referenziert wird.

<sup>630</sup> Der *Recall* ist die Abdeckung der relevanten Datenmenge durch die empfangenen Daten, während die in diesem Kontext häufig zitierte *Precision* den Anteil relevanter Daten in der empfangenen Datenmenge beschreibt (vgl. Mena et al. (2000): 412).

<sup>631</sup> Vgl. Abschnitt 4.3.3. Eine ähnliche Vorgehensweise im PDMS-Kontext zur Bestimmung einer ‚query degradation‘ relationaler Datenbankschemata ohne Berücksichtigung hierarchischer Strukturen verwenden auch FREIRE ET AL. (vgl. Freire et al. (2012): 174).

**Definition (22)** Die *Mapping-Qualität (Soundness)* ( $MQ_s$ ) ist definiert als:

$$MQ_s = \frac{1}{|D|} \sum_{d \in D} \frac{\text{Anzahl übertragbarer Elemente in } d}{\text{Anzahl aller Elemente in } d}, \text{ wobei } D \text{ die Menge der Dimensionen des Empfängers bezeichnet.}$$

Sie beschreibt damit eine Größe im Intervall  $[0; 1]$ , die ihren Maximalwert erreicht, wenn alle Elemente aller Dimensionen ein Gegenstück auf Seite des Datenliefernden Peers besitzen.

Die Betrachtung der Coherence (der Gleichartigkeit von Hierarchieebenen) kann mit der Betrachtung der Consistency zusammengefasst werden, da letztgenannte die Übereinstimmung der Hierarchieebenen bei der Betrachtung der Aggregatensebenen einzelner Dimensionselemente automatisch mit berücksichtigt.<sup>632</sup>

**Definition (23)** Die *Mapping-Qualität (Consistency)* ( $MQ_c$ ) ist definiert als:

$$MQ_c = \frac{1}{|D|} \sum_{d \in D} \frac{\text{Anzahl konsistent übertragener Elemente in } d}{\text{Anzahl übertragener Elemente in } d},$$

wobei ein Element  $m_i \in D$  konsistent ist, wenn gilt:

$$\mu(\rho_1(m_i)) = \rho_2(\mu(m_i)),^{633} \text{ d. h. das überführte Aggregations- (Eltern-)Element eines Elementes entspricht dem Aggregationselement seiner Überführung.}$$

Auch die  $MQ_c$  beschreibt eine Größe im Intervall  $[0; 1]$  und erreicht ihren Maximalwert, wenn alle Elemente, die übertragen werden, Aggregationskonsistent sind. Die  $MQ_c$  ist damit unabhängig von der Größe der  $MQ_s$  ermittelbar.<sup>634</sup>

Beide Größen stellen einen Mapping-spezifischen Wert dar, der jedoch die Spezifika der einzelnen Dimensionen nicht mehr abbilden kann. Zur Bildung einer MQ,

<sup>632</sup> Die Consistency ist somit eine detailliertere Form der Coherence. Die Definition beider begründet sich in der Verwendung als binärer Eigenschaften, sodass über die Mehrzahl der Eigenschaften eine Abstufung in der Dimensionsbetrachtung ermöglicht wird. Ein Mapping erreicht durch reine Übereinstimmung der Ebenen in den Hierarchien vergleichsweise schnell die Eigenschaft der Coherence, die Eigenschaft der Consistency ist hingegen durch die Betrachtung von erheblich mehr Elementen schwieriger zu erzielen.

<sup>633</sup> Vgl. Definition (2) und Definition (10). Die Schreibweise wurde durch  $m_i$  (im Vergleich zu einer Ausprägung von  $m(l_i)$ ) und  $\rho$  (im Vergleich zu  $\rho^{l \rightarrow l'}$ ) vereinfacht dargestellt.

<sup>634</sup> Sowohl  $MQ_s$  als auch  $MQ_c$  sind nicht definiert, wenn keine Elemente übertragen werden oder übertragen werden können. In diesem Fall kann jedoch auch gar kein Mapping existieren, sodass die MQ nicht bestimmt wird oder werden kann. Diese semantische Überführbarkeit ist daher als eine der Grundannahmen des CroCoBIN-Referenzmodells formuliert (vgl. Abschnitt 4.5.2.1). Ebenfalls als gegeben angenommen wird die von RIAZATI/THOM/ZHANG in diesem Zusammenhang definierte Strictness, deren Überprüfung somit entfällt (vgl. zur Strictness Definition (11); sowie für den Umgang mit nicht-strikten Dimensionen Abschnitt 4.3.4).

die die TMQ reflektiert, also auf die Auswertbarkeit einzelner Dimensionselementkombinationen Rücksicht nimmt, muss jedoch auch die Abbildung der einzelnen Dimensionen berücksichtigt werden. Abb. 47 stellt für ein Master-Schema (links) zwei mögliche Abdeckungen nach Überführung dar, die beide die gleiche TMQ aufweisen, jedoch unterschiedliche  $MQ_s$  (und bei entsprechender Hierarchie auch unterschiedliche  $MQ_c$ ),<sup>635</sup> da die Dimensionsstruktur, d. h. die Größe der Dimensionen, des Master-Schemas nicht berücksichtigt wird. Trotz der vermeintlich geringeren MQ ist Abdeckung (b) mindestens gleichwertig, je nach Analyseanspruch ggf. sogar als höherwertiger einzuordnen, da für alle betrachteten Dimensionen eine ähnliche Granularität aufrechterhalten wird.

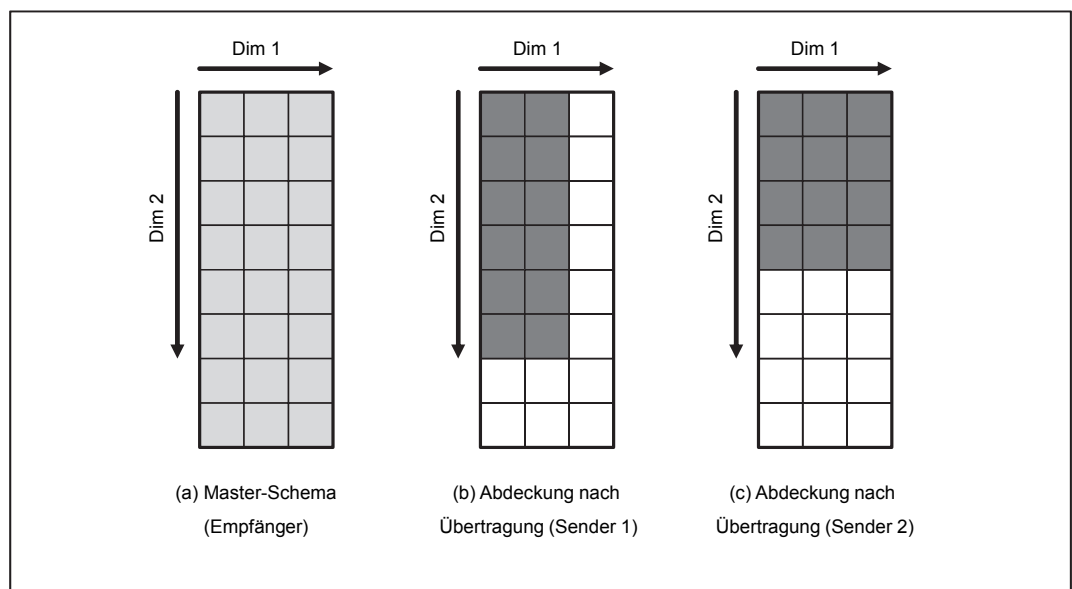


Abb. 47: Zwei TMQ-identische Abdeckungen eines Master-Schemas mit unterschiedlicher  $MQ_s/MQ_c$

(Quelle: Eigene Darstellung)

Werden die übertragenen Elemente je Dimension für jede Überführung in einem Vektor zusammengefasst, spiegelt dieser die gewichtete n-dimensionale Richtung wider, in der der Datenraum durch die Dimensionen aufgespannt wird.<sup>636</sup> Eine Überführung, deren Vektor dem des Master-Schemas ähnlich ist, ist daher mutmaßlich gut geeignet, dieses abzubilden. Abb. 48 stellt die unterschiedlichen Vektoren und ihre Winkel in Bezug auf den Vektor des Master-Schemas auf Basis des

<sup>635</sup> TMQ und  $MQ_s$  können im Beispiel einfach durch „Abzählen“ der Elemente bestimmt werden:

$$MQ_s(b) = \frac{1}{2} \left( \frac{2}{3} + \frac{6}{8} \right) = \frac{17}{24} < \frac{18}{24} = \frac{1}{2} \left( \frac{3}{3} + \frac{4}{8} \right) = MQ_s(c); TMQ(b) = TMQ(c) = \frac{12}{24} = 0,5.$$

<sup>636</sup> Vgl. Abschnitt 3.1.3.



gegebenen Beispiels dar. Es zeigt sich – an der im Gesamten um  $90^\circ$  gedrehten Ansicht – der deutliche kleinere Winkel  $\theta_b$  der Übertragung (b) aus Abb. 47.

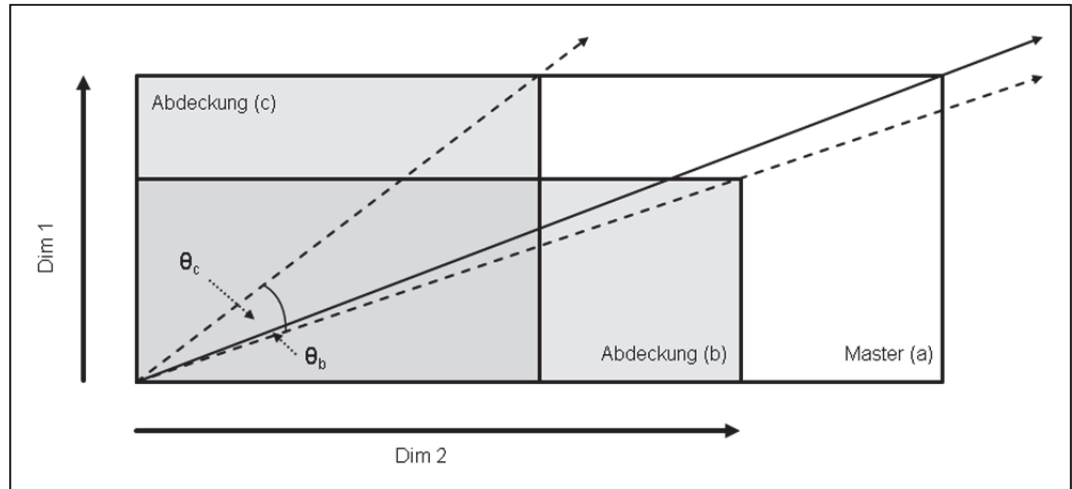


Abb. 48: Winkel zwischen Vektoren unterschiedlicher Abdeckungen gleicher TMQ und dem Vektor des Master-Schemas

(Quelle: Eigene Darstellung)

Zur Ähnlichkeitsbestimmung der Vektoren kann die – ebenfalls im IR gebräuchliche – *Kosinus-Ähnlichkeit* verwendet werden, deren Werte bei Vektoren mit ausschließlich nicht-negativen Komponenten im Intervall  $[0; 1]$  liegen, wobei der maximale Wert bei einer vollständig linearen Abhängigkeit der Vektoren gegeben ist.<sup>637</sup> Es gilt für die Kosinus-Ähnlichkeit  $\cos \theta$  zweier Vektoren  $a, b$ :<sup>638</sup>

**Definition (24)**  $\cos(\theta) = \frac{a \cdot b}{\|a\| \|b\|}$ , wobei  $\|a\|$  die euklidische Norm beschreibt:

$$\|a\| = \sqrt{\sum_{i=1}^n (a_i)^2}$$

Es ergibt sich entsprechend:

**Definition (25)** Die *Mapping-Qualität (Balance)* ( $MQ_b$ ) ist definiert als:

$$MQ_b = \cos(\theta), \text{ mit}$$

$\theta$  als Winkel zwischen  $a$  und  $b$ , wobei

$$a = (|\mu(d_1)|, \dots, |\mu(d_i)|, \dots, |\mu(d_{|D|})|) \text{ und}$$

$$b = (|d_1|, \dots, |d_i|, \dots, |d_{|D|}|), \text{ wobei}$$

<sup>637</sup> Vgl. Singhal (2001): 36f.

<sup>638</sup> Vgl. Nguyen/Bai (2011): 711.



$|d_i|$  die Anzahl der Elemente in Dimension  $i$  und  $|\mu(d_i)|$  die Anzahl der für Dimension  $i$  überführbaren Elemente abbilden (jeweils aus Sicht des Masters).

Die Teilmaße  $MQ_s$ ,  $MQ_c$  und  $MQ_b$  können dann mit den einzelnen Faktoren gewichtet werden und ergeben in ihrer Kombination die eigentliche MQ:<sup>639</sup>

**Definition (26)**  $MQ = \frac{\alpha MQ_s + \beta MQ_c + \gamma MQ_b}{\alpha + \beta + \gamma}$ , mit  $\alpha, \beta, \gamma \in \mathbb{R}^{>0}$

Die MQ ist dabei immer als Messgröße einer gerichteten Verbindung aus Sicht des Empfängers zu interpretieren. Die Notation  $MQ(v_1, v_2)$  – analog  $MQ(e_{12})$ , mit  $e_{12} = (v_1, v_2)$  – bezeichnet daher die MQ, die bei der Übertragung von  $v_2$  an  $v_1$  erzielt wird.<sup>640</sup>

### 5.3.2 Pfad-Qualität

Zur Bestimmung der Qualität eines Pfades, also der Kombination mehrerer Verbindungen, die gerichtet von einem Peer zu einem anderen führen, müssen die Qualitäten der einzelnen Verbindungen kombiniert werden. Die Kontrollgröße für die PQ ist die *tatsächliche Pfad-Qualität* (TPQ), die sich ermitteln lässt, wenn bei vollständiger Kenntnis der Schemata und Überleitungen aller beteiligten Peers geprüft wird, welche Daten übermittelt werden können. Abb. 49 zeigt ein einfaches Pfad-Beispiel, bei dem die TPQ durch Abzählen der übertragbaren Elemente, d. h. Datenzellen (gekennzeichnet durch Zahlen), ermittelt werden kann. Die  $TPQ(v_1, v_3)$ <sup>641</sup> entspricht dann den Elementen aus  $v_3$  die sowohl über  $(v_3, v_2)$  als auch über  $(v_2, v_1)$  geführt werden (im Beispiel die Elemente 2 und 3) im Verhältnis zu allen Elementen von  $v_1$  und ist damit analog der TMQ beschrieben.<sup>642</sup>

<sup>639</sup> Zur Bestimmung sinnvoll verwendbarer Parameterwerte der Gewichtungsfaktoren im gegebenen Kontext vgl. Abschnitt 6.4.1.

<sup>640</sup> Im der vorliegenden Arbeit wird der Plural der Qualitäten in gleicher Abkürzung verwendet (‘MQ’, nicht ‘MQs’), da – anders als bei englischsprachigen Ausdrücken – auch die ausgeschriebene Form kein Plural-s erhält. Es wird so zudem eine Verwechslung mit der  $MQ_s$  vermieden.

<sup>641</sup> Die Notation wird analog der MQ-Notation geführt (vgl. Abschnitt 5.3.1).

<sup>642</sup> Zu beachten ist hierbei, dass die TPQ – wie auch die TMQ – die potenziell übertragbaren Daten betrachtet. Es wird keine Aussage darüber getroffen, ob  $v_3$  tatsächlich Informationen – im Sinne von Fakten – in allen durch das Schema und die Überführungen beschriebenen Datenzellen vorhält.

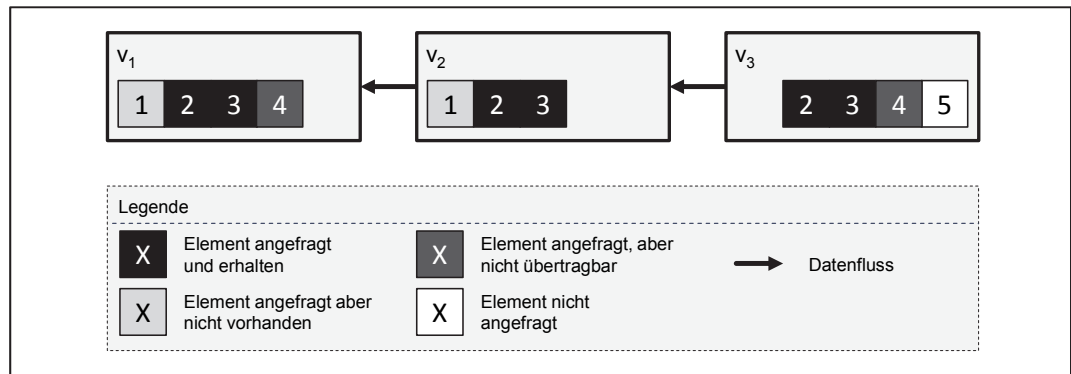


Abb. 49: Darstellung eines einfachen Pfades zur Ermittlung der TPQ

(Quelle: Eigene Darstellung)

Für die Verknüpfung der MQ der Verbindungen im Pfad muss eine Funktion  $t$  verwendet werden, die die folgenden Eigenschaften erfüllt:

1. Da MQ im Intervall  $[0; 1]$  liegt, muss gelten:  $t: [0; 1] \times [0; 1] \rightarrow [0; 1]$ .
2. Da ein perfektes Matching die PQ nicht verringern darf, ist 1 ein neutrales Element, d. h.  $t(x, 1) = t(1, x) = x$ .
3. Da bessere MQ zu besseren PQ führen sollen, muss  $t$  monoton steigend sein, d. h.  $t(x, y) \leq t(x', y')$ , wenn  $x \leq x'$  und  $y \leq y'$ .
4. Die Reihenfolge der Peers, über die der Pfad läuft, muss unerheblich sein, da nur die MQ der Verbindungen berücksichtigt werden sollen, d. h. es muss Kommutativität vorliegen, d. h.  $t(x, y) = t(y, x)$ .<sup>643</sup>
5. Wird eine Verbindung durch Einfügen eines neuen Peers verlängert, so muss die Stelle des Einfügens unerheblich sein, solange die MQ zu den jeweiligen Nachbarpeers identisch ist, d. h. es muss Assoziativität vorliegen, d. h.  $t(x, t(y, z)) = t(t(x, y), z)$ .

Eine einfache Funktion, die diese Anforderungen erfüllt, ist das algebraische Produkt, das daher auch in bestehenden Ansätzen verwendet wird.<sup>644</sup> Die gestellten Anforderungen werden jedoch von einer Reihe von Funktionsfamilien erfüllt, die als *Triangular-Normen* oder *t-Normen* beschrieben sind und zu denen diverse Konstruktionsvorschläge existieren.<sup>645</sup> Sie unterscheiden sich dabei deutlich hin-

<sup>643</sup> Die Kommutativität der Verknüpfungsfunktion ist dabei nicht zu verwechseln mit der Kommutativität der MQ eines einzelnen Matchings (vgl. Fußnote 627).

<sup>644</sup> Vgl. Mandreoli et al. (2006): 21.

<sup>645</sup> Vgl. Fagin (1999): 85; Klement/Mesiar/Pap (2005): 17–19.

sichtlich ihrer Ausprägung und der Geschwindigkeit ihrer Berechnung.<sup>646</sup> Abb. 50 zeigt dreidimensionale Plots mehrerer t-Normen unterschiedlicher Familien.<sup>647</sup> Es lässt sich direkt erkennen, dass dabei die Minimum-Norm  $T_M$  die stärkste darstellt, d. h. es gilt für alle t-Normen  $T_i$ :  $T_i(x, y) \leq T_M(x, y)$ , mit  $(x, y) \in [0,1]^2$ .<sup>648</sup> Ebenfalls ersichtlich ist der große Unterschied in der Beschaffenheit des Wertebereichs. Eine Konstruktionsübersicht der Normen gibt im Anschluss Tab. 11.

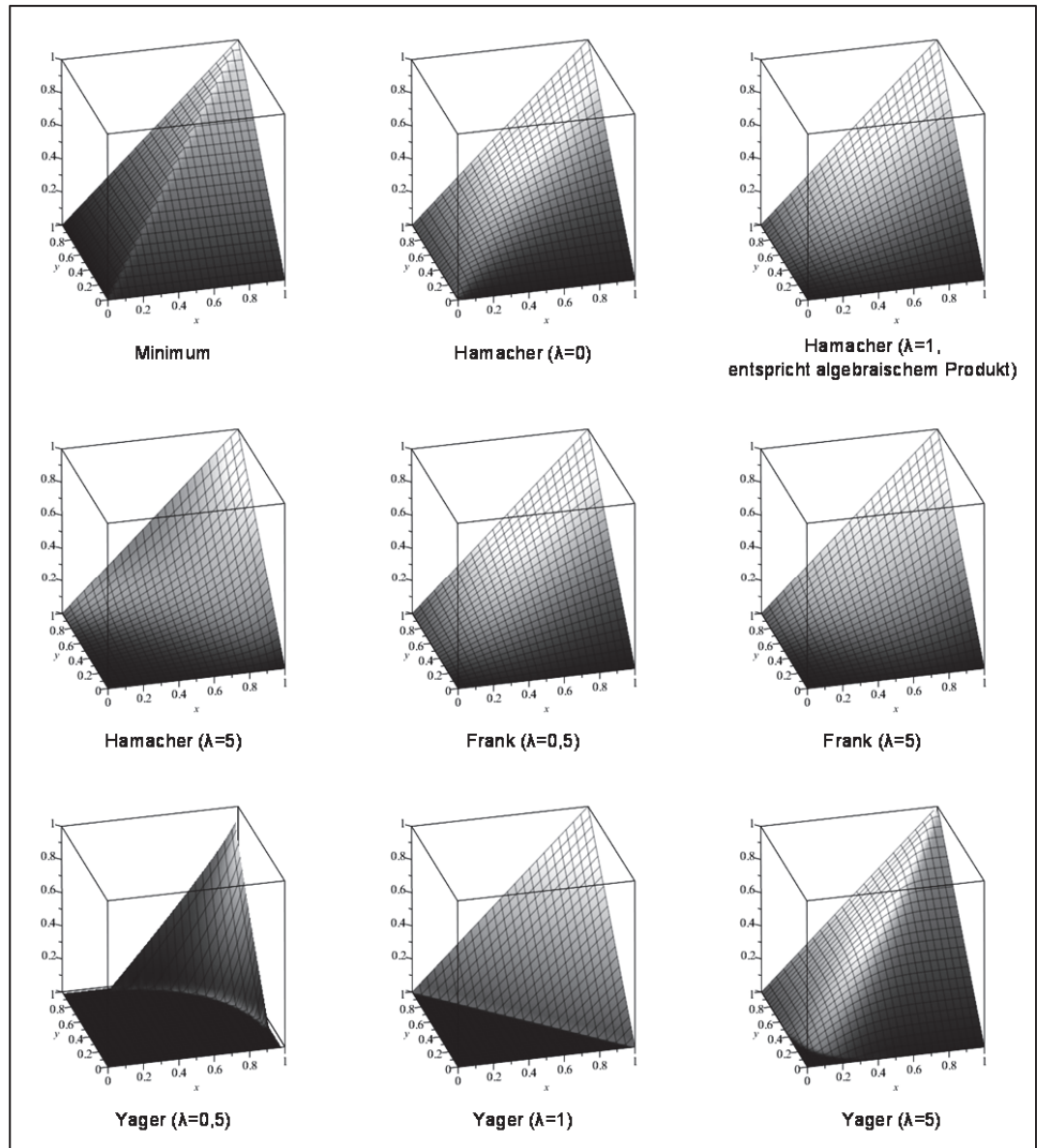


Abb. 50: Plots unterschiedlicher t-Normen

(Quelle: Eigene Darstellung)

<sup>646</sup> Vgl. Eckhardt/Skopál/Vojtáš (2009): 70f.

<sup>647</sup> Die Auswahl der Familien und Parameter erfolgt dabei mit der Absicht, die Heterogenität der Normen aufzuzeigen (siehe z. B. Yager ( $\lambda=0,5$ ) und Minimum) und vor dem Hintergrund der Anwendung in der prototypischen Implementierung (vgl. Abschnitte 6.3.2 und 6.4.2).

<sup>648</sup> Vgl. Klement/Mesiar/Pap (2005): 19.

Name	Definition
<b>Minimum</b>	$T_M(x, y) = \min(x, y)$
<b>Hamacher</b>	$T_\lambda^H(x, y) = \begin{cases} 0 & , \text{für } \lambda = x = y = 0 \\ \frac{xy}{\lambda + (1-\lambda)(x+y-xy)} & , \text{für } \lambda \in ]0; \infty[ \end{cases}$
<b>Frank</b>	$T_\lambda^F(x, y) = \begin{cases} T_M(x, y) & , \text{für } \lambda = 0 \\ T_1^H(x, y) & , \text{für } \lambda = 1 \\ \log_\lambda \left( 1 + \frac{(\lambda^x - 1)(\lambda^y - 1)}{\lambda - 1} \right) & , \text{für } \lambda \in ]0; 1[ \cup ]1; \infty[ \end{cases}$
<b>Yager</b>	$T_\lambda^Y(x, y) = \max(1 - ((1-x)^\lambda + (1-y)^\lambda)^{\frac{1}{\lambda}}, 0)$ , für $\lambda \in ]0; \infty[$

Tab. 11: Konstruktionsübersicht unterschiedlicher t-Normen

(Quelle: Eigene Darstellung)

Die Wahl einer geeigneten t-Norm hat hohen Einfluss auf die Ergebnisse in der Netzgestaltung, da bereits t-Normen der gleichen Familie für identische Verbindungen nicht nur unterschiedliche Werte, sondern auch unterschiedliche Bewertungsreihenfolgen hervorbringen können. Tab. 12 stellt dies beispielhaft für zwei Normen der ‚Hamacher‘-Familie dar – Hamacher ( $\lambda=1$ ) entspricht dabei dem algebraischen Produkt.

Verbindung A	Verbindung B	Hamacher ( $\lambda=1$ )	Hamacher ( $\lambda=5$ )
<b>MQ = 0,7</b>	<b>MQ = 0,4</b>	0,28 (Rang 1)	~0,16 (Rang 2)
<b>MQ = 0,9</b>	<b>MQ = 0,3</b>	0,27 (Rang 2)	~0,21 (Rang 1)

Tab. 12: Bewertungsunterschiede bei Einsatz unterschiedlicher t-Normen

(Quelle: Eigene Darstellung)

Die Auswahl der t-Norm kann daher nicht ohne Prüfung des Einsatzgebietes bestimmt werden, sondern wird im Rahmen dieser Arbeit für den Einsatz in einem CroCoBIN bestimmt.<sup>649</sup> Für die reine Beschreibung der PQ ist dies allerdings zunächst unerheblich, sodass für diese die folgende Definition gilt:

**Definition (27)**  $PQ = t(MQ(v_0, v_1), \dots, MQ(v_i, v_{i+1}), \dots, MQ(v_{n-1}, v_n))$ , mit  $P(v_0, v_n)$  als betrachtetem Pfad und  $t(x_1, \dots, x_n) = t(x_1, t(x_2, t(\dots, x_n)))$ , wobei  $t$  eine t-Norm ist.

<sup>649</sup> Vgl. Abschnitt 6.4.2.

Analog zur MQ ist die PQ Messgröße gerichteter Verbindungen aus Sicht des Empfängers. Die Notation  $PQ(v_1, v_2)$  bezeichnet daher die PQ, die bei der Übertragung von  $v_2$  an  $v_1$  über alle Knoten des Pfades erzielt wird.

### 5.3.3 Routing-Qualität

Die Bestimmung der PQ ermöglicht es, bei der Auswahl von Routings nicht nur die Überführungsqualität zu den direkten Nachbarn mit zu berücksichtigen. Abb. 51 stellt ein Netzwerk mit fünf Teilnehmern und – ausgehend von  $v_1$  – den nutzbaren Pfaden dar.

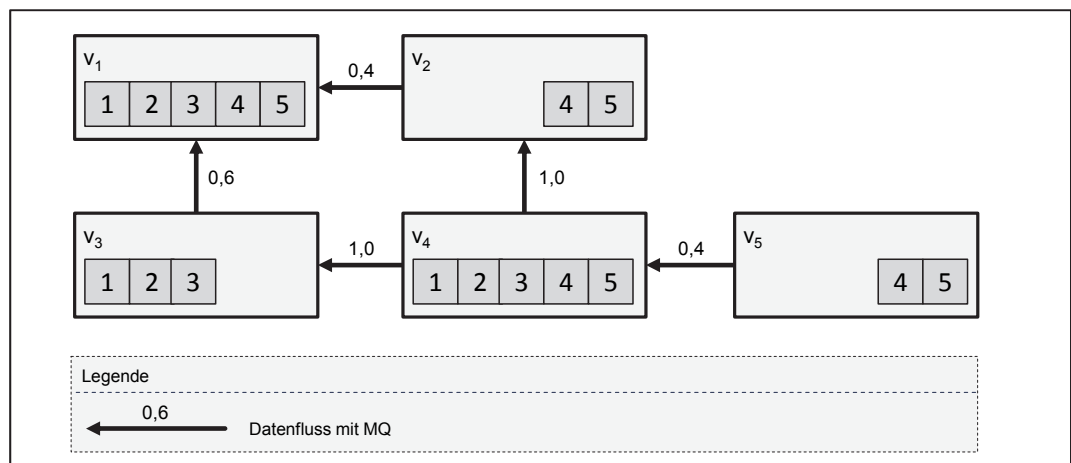


Abb. 51: Alternative Routing-Möglichkeiten in einem 5-elementigen Netzwerk  
(Quelle: Eigene Darstellung)

Wenn  $v_1$  eine Abfrage an das Netzwerk stellt, würde diese direkt an die Nachbarn  $v_2$  und  $v_3$  gerichtet, welche dann die Abfrage weiterleiten können. Dabei ist festzulegen, auf welcher Grundlage entschieden wird, wessen Abfrage tatsächlich bei  $v_4$  eingeht, da  $MQ(v_3, v_4) = MQ(v_2, v_4) = 1,0$ . Sofern  $v_1$  auf Basis seines zunächst auf die Nachbarschaft begrenzten Wissens eine Weiterleitung über  $v_3$  (wegen der höheren MQ) präferiert, ergibt sich ein Routing  $R_1$ , das neben den direkten Verbindungen die Pfade  $(e_{13}, e_{34})$  und  $(e_{13}, e_{34}, e_{45})$  enthält. Es lässt sich im Beispiel durch einfaches Abzählen dabei eine Anzahl von acht empfangenen Elementen feststellen, während das alternative – aber zunächst weniger attraktiv erscheinende – Routing  $R_2$  über  $v_2$  zu neun empfangenen Elementen führt. Die Be-

rücksichtigung der tatsächlichen Pfadqualitäten zu allen Knoten im Netzwerk kann somit zu erheblich besseren Ergebnissen führen.<sup>650</sup>

Um dieses Wissen nutzen zu können, ist eine *statische* oder *zentrale* Routing-Strategie erforderlich, bei der jeder Knoten über Informationen über den besten Pfad zu jedem anderen Knoten im Netzwerk verfügt und diesen aktiv wählt, d. h. eine Abfrage gezielt an einen Nachbarknoten stellt mit der Information, an wen diese weiterzureichen ist.<sup>651</sup> Trotz des Charakters der Unabhängigkeit der Beteiligten in CroCoBIN-basierten Netzwerken, sind die erwartete Fluktuation und Netzgröße auf Basis der grundsätzlich angenommenen mittel- bis langfristigen Zusammenarbeit im Vergleich zu aufwändigen Netzwerken wie dem Internet so gering, dass der Aufwand zur Erstellung der Tabelle angemessen erscheint.<sup>652</sup>

Die Auswahl des jeweils bestmöglichen Pfades (auf Basis der MQ benachbarter Knoten) zu jedem Knoten im Netzwerk bestimmt dann für einen Knoten  $v$  die Qualität seines Routings als die durchschnittliche Qualität aller verwendeten Pfade, die angibt, mit welcher Qualität im Mittel bei einer Abfrage an das gesamte Netzwerk gerechnet werden kann:

**Definition (28)**  $RQ = \frac{1}{n-1} \sum_{i=2}^n PQ(v_1, v_i)$ , mit

$v_1$  als betrachtetem Knoten und

$n$  Teilnehmern im Netz.

Das Ausrichten des Routings an den Qualitäten im Netzwerk (und nicht an den Pfadlängen) wird – in Anlehnung an entsprechende Ansätze im P2P-Fileharing-

<sup>650</sup> Während im bereits mehrfach als Vergleich herangezogenen PDMS-Kontext FREIRE ET AL. im Rahmen moderner Routing-Strategien zwar IQ-Kriterien nach WANG/STRONG berücksichtigen, wenden sie diese jedoch ausschließlich als lokal begrenztes Wissen an. Eine Betrachtung kompletter Pfade findet in diesem Kontext nicht statt (vgl. Freire et al. (2013): 311). Das beschriebene Vorgehen ähnelt durch das lokal begrenzte Wissen und die zunächst egoistische, aber nicht böswillige Handlung, den ‚besten‘ Nachbarn zu bedienen, dem von ROUGHGARDEN/TARDOS beschriebenen ‚selfish routing‘ (vgl. Roughgarden/Tardos (2002): 237).

<sup>651</sup> Es existiert eine Vielzahl unterschiedlicher Routing-Strategien, die lokales oder globales Wissen miteinbeziehen und die Routing-Koordination darauf aufbauend implementieren. Für eine ausführliche Einführung in die Thematik – und tiefergehende Betrachtungen – vgl. Medhi/Ramasamy (2007).

<sup>652</sup> Zum Vergleich zu einem CroCoBIN mit typischerweise wenigen dutzend bis einigen hundert Teilnehmern, umfasst die Routing-Tabelle der Hauptknotenpunkte des Internets derzeit mehr als 500.000 Einträge (vgl. Shen (2014)), ohne dass eine mehr als technische, leicht austauschbare Kopplung dieser vorhanden ist – im Gegensatz zu den Teilnehmern eines CroCoBIN, wo ein einmal geschaffenes Mapping auf Grund des hohen Erstellungsaufwandes nicht einfach ersetzt werden kann.

Bereich –<sup>653</sup> als ‚semantic routing‘ bezeichnet. Der RQ kann dabei – analog zu TMQ und TPQ – als Kontrollgröße die *tatsächliche Routing-Qualität* (TRQ) gegenübergestellt werden, die – analog zur RQ – als Durchschnitt der TPQ aller verwendeten Pfade bestimmt wird.<sup>654</sup>

### 5.3.4 Netz-Qualität

Die bisherigen Überlegungen zu Qualitäten beziehen sich auf die Güte der Überführung von Daten zwischen Netzteilnehmern. Wird ein Netzwerk in Summe betrachtet, lassen sich weitere Eigenschaften identifizieren, die sich z. B. aus seiner Topologie ergeben. Ein Hauptbetrachtungsgegenstand ist dabei die Robustheit des Netzes, also die Aufrechterhaltung der Funktionsfähigkeit bei Entfernung von Knoten. Die Robustheit steigt mit der relativen Verbindungsanzahl eines Netzwerks, d. h. dessen Dichte. Die Größe eines Netzwerks hingegen nimmt weniger Einfluss auf die Robustheit, d. h. große und kleine Netzwerke – sowohl Zufallsnetzwerke als auch skalenfreie Netzwerke – verhalten sich ähnlich bei der Entfernung einer bestimmten Prozentzahl von Knoten aus einem existierenden Netz.<sup>655</sup> Als entscheidender Faktor für die Robustheit lässt sich vielmehr die Auswahl der Grundtopologie identifizieren.<sup>656</sup>

Zur Messung der Robustheit existiert kein übergreifendes Maß. Es werden aber diverse einzelne Messgrößen vorgeschlagen, die sich in klassische, aus der Graphentheorie abgeleitete, und dynamische Größen unterscheiden und nicht nur die reine Topologie betrachten, sondern auch z. B. Gruppenbildungen oder Gewichte

<sup>653</sup> Vgl. Chehade et al. (2009): 1.

<sup>654</sup> Die in Abb. 51 vorgestellte Problematik bezieht bei der Beschreibung die TRQ bereits implizit mit ein, da das schlechtere Ergebnis bei Verwendung von Routing  $R_1$  bereits die tatsächlich zu erzielende Qualität widerspiegelt. Die Verwendung der PQ unterliegt – wie auch die der MQ – den bereits diskutierten Restriktionen eines nicht-vollständigen Wissens über das Netzwerk und ist ‚nur‘ dazu geeignet, die beschriebene Problematik – in diversen Ausprägungen – auf Basis des vorhandenen Wissens zu verringern, nicht aber ein im Sinne der TRQ ‚optimales‘ Routing zu erstellen. Im Vorgriff auf die Implikationen auf die weitere Forschung (vgl. Abschnitt 7.2) sei hier bereits die Idee skizziert, die Qualitäten dimensionsspezifisch zu hinterlegen, sodass Abfragen an Peers im Netzwerk aufgeteilt werden können, um hohe Granularitäten in einzelnen Dimensionen bei einigen Peers nutzbar zu machen. In einem CroCoBIN ist das auf Grund der Ein-Wege-Bedingung zur Aufrechterhaltung der Aggregationsintegrität zunächst nicht möglich (vgl. Abschnitt 4.5.2.4), könnte aber bei anderweitiger Sicherstellung dieser realisierbar werden. Vergleichbar sind Überlegungen zur Aufteilung von Datenbank-Abfragen in verteilten Systemen (vgl. Naumann/Leser/Freytag (1999): 447).

<sup>655</sup> Vgl. Chen/Li/Ke (2012): 524.

<sup>656</sup> Vgl. Albert/Barabási (2002): 87; Jamakovic/Uhlig (2007): 283; sowie die Ausführungen in Abschnitt 5.2.2.



der Kanten.<sup>657</sup> Dabei zeigen sich die klassischen Maße im Vergleich weniger aussagefähig als spezialisierte dynamische Größen, bieten aber in Kombination dennoch einen guten Einblick in das Netzverhalten.<sup>658</sup> Allen Messgrößen gemein ist, dass sie nicht normiert sind – und in erster Linie dem Vergleich von Netzen gleicher Art dienen – sowie nicht den statischen Zustand eines Netzes beschreiben können, sondern die Netzwerkgröße bei Entfernen einer bestimmten Anzahl Knoten beschreiben.<sup>659</sup> Sie sind zudem teilweise nur sehr aufwändig ermittelbar.<sup>660</sup>

Für die vorliegende Arbeit wird daher auf den quantifizierten Einbezug eines Robustheitsmaßes in die übergreifende NQ verzichtet. Die durchgeführten Simulationen und deren Auswertungen weisen aber zentrale Standardmaße der Netzwerke aus. Zu diesen gehören der *Clustering Coefficient* (CC) und die *durchschnittliche Pfadlänge* (*Average Path Length* (AvgPL)).<sup>661</sup> Der CC bestimmt dabei lokal, wie stark die Nachbarschaft eines beliebigen Knotens miteinander verbunden ist. Seine globale Definition ist der Mittelwert der lokalen Koeffizienten:<sup>662</sup>

**Definition (29)**  $CC_{global} = \frac{1}{n} \sum_{i=1}^n \frac{2|E_i^N|}{|N_i| \cdot (|N_i| - 1)}$ , wobei

$n$  die Zahl der Knoten im Netz,

$N_i$  die Nachbarschaft des Knotens  $i$  und

$E_i^N$  die Anzahl existierender Verbindungen innerhalb der Nachbarschaft eines Knoten  $i$  beschreibt.

Die vorliegende Definition hat sich im Rahmen von Computersimulationen aufgrund ihrer einfachen Berechenbarkeit als Standard herausgebildet. Es existieren zum CC in der Literatur jedoch alternative Vorschläge. Bei der Interpretation der

<sup>657</sup> Vgl. Manzano et al. (2013): 3643; sowie für eine weitere ausführliche Übersicht diverser Messgrößen, ihrer Klassifikationsmöglichkeiten und einzelner Anwendungstests Grubisic et al. (2008): 88–112.

<sup>658</sup> Vgl. Manzano/Calle/Harle (2011): 5–7.

<sup>659</sup> Vgl. Manzano/Calle/Harle (2011): 2f.

<sup>660</sup> Ein Beispiel dafür ist der von JAMAKOVIC/UHLIG für die Analyse vorgeschlagene zweitkleinste Eigenwert einer Matrix der Größe  $n \times n$  ( $n$  ist die Größe des Netzwerks), der für jeden Knoten errechnet werden muss (vgl. Jamakovic/Uhlig (2007): 277).

<sup>661</sup> Vgl. Albert/Barabási (2002): 49.

<sup>662</sup> Vgl. Watts/Strogatz (1998): 441.



Kennzahl muss daher immer auf die entsprechende Definition referenziert werden.<sup>663</sup>

Neben der Robustheit können weitere Eigenschaften der Netze getestet werden. So ist z. B. die Sicherstellung einer ausgewogenen Belastung der Netzteilnehmer denkbar, damit qualitativ hochwertige Knoten nicht zu technisch – und damit ggf. auch finanziell – überbeanspruchten Netzteilnehmern werden, was bspw. in einem Stern-topologisch ausgerichteten Netzwerk der Fall ist. Derartige Überlegungen übersteigen jedoch die zunächst an fachlichen Zielen ausgerichtete Betrachtung innerhalb der vorliegenden Arbeit und das angewandte CroCoBIN-Referenzmodell.<sup>664</sup>

Als übergeordnete Kennzahl für die Qualität eines Netzwerks wird daher aufbauend auf den bisherigen Qualitäts-orientierten Kennzahlen die NQ als Mittel der RQ aller im Netzwerk vorhandenen Peers definiert. Sie gibt damit die durchschnittliche erzielte Qualität der Antwort auf eine von einem beliebigen Peer gestellte Abfrage wieder und kann – analog der RQ und TRQ – mit einer *tatsächlichen Netzqualität* (TNQ) als Kontrollgröße versehen werden, die die tatsächliche, durchschnittliche Antwortqualität abbildet.

**Definition (30)**  $NQ = \frac{1}{n} \sum_{i=1}^n RQ(v_i)$ , mit

$v_i \in V$  als betrachtetem Knoten und  
 $n$  Teilnehmern im Netz.

Die NQ ist damit abhängig von der Ausprägung der Gewichtungsfaktoren der MQ und der gewählten t-Norm für die Verknüpfung von Verbindungsqualitäten sowie – durch die unterschiedlich häufige Anwendung der t-Norm bei kleinen und großen Pfadlängen – von der Netzgröße. Bei gleichen Gewichtungsfaktoren besteht in bestimmten Konstellationen die Möglichkeit, zumindest annäherungsweise eine Vergleichbarkeit von Netzqualitäten herzustellen, indem eine Umkehrfunktion auf die t-Norm angewendet wird und so die durchschnittliche MQ der Verbindungen in den Pfaden getestet wird.<sup>665</sup> Dieses Vorgehen ist aber z. B. für die Minimum-

<sup>663</sup> Vgl. Newman (2003): 12.

<sup>664</sup> Vgl. Abschnitte 4.5.2 und 5.2.1.

<sup>665</sup> So ergibt sich z. B. für die Anwendung des algebraischen Produktes die Möglichkeit, die  $n$ -te Wurzel des Endwertes zu bestimmen, wobei  $n = AvgPL - 1$  die Anzahl der durchschnittlichen Anwendungen der t-Norm beschreibt.

oder Yager-Normen aufgrund der nicht eindeutigen Umkehrbarkeit nicht möglich, sodass ein Vergleich der durchschnittlichen MQ erheblich einfacher ist und – bei Anwendung einer MQ-orientierten Netzaufbaustrategie – zur NQ ähnliche Resultate hervorbringt.<sup>666</sup>

#### 5.4 Einflussfaktoren auf die Güte des Netzes

Die bisher vorgestellten Einflussfaktoren für Messgrößen wirken sich zunächst auf die Messgrößen selbst, bei Verwendung dieser als Entscheidungskriterium – bei der Pfadauswahl für das Routing – aber auch indirekt auf die tatsächliche Güte des Netzwerks aus. Außerhalb dieser Messgrößen-orientierten Betrachtung lassen sich zudem weitere Einflussfaktoren auf die Güte des Netzes identifizieren.<sup>667</sup>

Im Kern der Überlegungen steht dabei stets die Nachbarschaftswahl, d. h. die Bestimmung für einen (zumeist neu in das Netzwerk eintretenden) Knoten, zu welchen weiteren Knoten Verbindungen hergestellt, also Mappings geschaffen werden sollen. Ohne tiefere Betrachtung lassen sich diverse Ideen rechtfertigen, die z. B. die ähnlichsten Knoten, die umfangreichsten Knoten oder mehrere, im Netzwerk voneinander entfernte Knoten wählen, um die Übertragungsqualität zu den direkten Nachbarn oder der Netzgesamtheit positiv zu beeinflussen. Ähnliche Fragestellungen und Ansätze ergeben sich bei der Gestaltung von P2P-Netzwerken zum Datei- und Dokumentenaustausch. Dort existiert unter anderem – auch aufgrund der wesentlich höheren Teilnehmerzahl – der Vorschlag, grundsätzlich nicht das gesamte Netzwerk zu berücksichtigen, sondern ein Teilnetzwerk zu identifizieren, das eine hohe semantische Ähnlichkeit besitzt, sodass für zukünftige Abfragen – die auch nur an dieses Teilnetzwerk gestellt werden – eine hohe Wahrscheinlichkeit für die vollständige Beantwortung dieser besteht. Dieses Konzept eines „Semantic Overlay Network“<sup>668</sup> führt allerdings zu einer bewussten Inkaufnahme von Unvollständigkeit bei der Abfrageerstellung und der Fokussierung auf Pfadlängen (statt auf Qualitäten).<sup>669</sup> Es ist damit nicht direkt auf CroCoBIN-basierte Netzwerke übertragbar.

---

<sup>666</sup> Vgl. Abschnitt 6.4.

<sup>667</sup> Vgl. Abschnitt 5.2.1.

<sup>668</sup> Crespo/Garcia-Molina (2005): 1.

<sup>669</sup> Vgl. Lodi et al. (2008): 26–28, 33; sowie die Hinweise zu Superpeer-Schemata in Abschnitt 4.3.4.

Ebenfalls ungeeignet sind die naiven Ansätze, als Einstiegspunkte Knoten mit einer hohen RQ zu wählen, um – nach Anwendung der t-Norm auf die zusätzliche Verbindung zu diesen Knoten – selbst eine hohe RQ zu erzielen, oder die eigene potenzielle RQ zunächst auf Basis der bestehenden Einstiegspunkte-RQ zu errechnen, um diese dann zu maximieren. Zum einen können, selbst wenn die RQ aller Knoten vollständig publiziert sind, die Überführungswerte aufgrund fehlender Mappings nur approximiert werden, zum anderen lässt sich mit einem Knoten, der in allen im Netzwerk verfügbaren Dimensionen lediglich ein hochaggregiertes Element vorhält, ein einfaches Beispiel für einen Peer konstruieren, der – weil er sehr wenig Daten anfragt und diese also häufig erhalten wird – eine sehr hohe RQ besitzt, aber keinen guten Einstieg in das Netzwerk bedeutet, da jede Abfrage, die über ihn läuft, nahezu alle Daten der ‚vorher‘ liegenden Peers verlieren wird.

Die Nachbarschaftswahl kann sich daher nur an der ‚lokalen‘ MQ zwischen dem neu eintretenden Peer und den vorhandenen Peers orientieren sowie Rahmenparameter des Netzes ausrichten, die sich z. B. aus einer Vorgabe an Verbindungsanzahlen oder dem Konstruktionsalgorithmus zur Erstellung einer bestimmten Netztopologie ergeben.<sup>670</sup> Die MQ ist dabei nicht ex ante bestimmbar, ihr kann sich aber durch automatisierte Verfahren angenähert werden.<sup>671</sup> Die dadurch entstehende *prognostizierte Mapping-Qualität* (PMQ) ermöglicht – je nach Güte – eine qualifizierte Auswahl.

Eine solche Nachbarschaftswahl auf der Basis der jeweils besten verfügbaren MQ entspricht im Wesentlichen dem Vorgehen eines ‚gierigen‘ (engl.: ‚greedy‘) Algorithmus. Ein solcher wählt als nächstes zu verarbeitendes Element immer das nach einem gegebenen Kriterium beste.<sup>672</sup> Dieses Vorgehen unterliegt jedoch ähnlichen Einschränkungen wie denen bei der Auswahl des Nachbarknotens auf Basis dessen RQ. Während ein ‚selbstsüchtiges‘ Verfahren dort zu einer problematischen Nachbarschaftswahl führen kann, ist bei der Verwendung der MQ ein ‚selbstloses‘ Vorgehen fehleranfällig, da die MQ in Richtung des potenziellen (Nachbar-

---

<sup>670</sup> Vgl. Abschnitt 4.5.2.

<sup>671</sup> Vgl. Abschnitt 5.2.1; sowie insbesondere Fußnote 588.

<sup>672</sup> Vgl. Neapolitan (2014): 151.

schaft-)Zielknotens ggf. nur deshalb hoch ist, weil der Zielknoten ein stark reduziertes Schema aufweist.<sup>673</sup>

Für die Rahmenbedingungen lassen sich umfangreiche Parameter beschreiben (z. B. Größe des Netzwerks, Anzahl der Verbindungen, Zielverbindungsgrad, Ziel-Topologie, Größe eines Basisnetzwerks mit vollständiger Verbindung), die teilweise abhängig voneinander wirken. Dabei können naheliegende Zusammenhänge vermutet werden („Je dichter das Netzwerk, desto besser die NQ.“), eine zuverlässige Aussage darüber ist aber nur auf Grundlage einer umfangreichen Datenbasis möglich. Das folgende Kapitel stellt daher eine prototypische Implementierung eines CroCoBIN-basierten Netzwerks vor, die als Simulationswerkzeug dient, um eine Vielzahl von Parameterkombinationen und deren Auswirkungen auf die Netzwerke und ihre Qualität darzustellen und auszuwerten.<sup>674</sup>

---

<sup>673</sup> Die Unterschiede beider Verfahren und die daraus folgenden Implikationen auf die Qualitäten des Netzes werden ausführlich bei der Betrachtung der Ergebnisse der Simulation dargestellt (vgl. Abschnitt 6.4).

<sup>674</sup> Für eine Darstellung der beschriebenen Parameter und ihrer Ausprägungsmöglichkeiten vgl. insbesondere Abschnitte 6.3.2 und 6.3.2.2.

## 6 Bestimmung sinnvoller Konfigurationen von Business-Intelligence-Netzwerken

In den vorhergehenden Kapiteln wurde ein Referenzmodell für ein P2P-BI-Netzwerk vorgestellt (CroCoBIN) und anhand dessen Einflussfaktoren, Rahmenbedingungen und Gütemaße entwickelt.<sup>675</sup> Diese können dazu genutzt werden, Erkenntnisse über die beschriebenen Netzwerke zu erlangen und Empfehlungen zu entwickeln, welche Eigenschaften und Vorgaben vorteilhaft für die Implementierung von unternehmensübergreifenden BI-Netzwerken in der Praxis sind. Da derartige Netzwerke weder ausreichend erforscht noch tatsächlich vorhanden sind,<sup>676</sup> wird auf Modelle zurückgegriffen, anhand derer unterschiedliche Netzkonfigurationen betrachtet werden können.

Das Durchführen entsprechender Experimente und Auswertungen wird im Rahmen der vorliegenden Arbeit als *Simulation* bezeichnet. Der Begriff der Simulation ist dabei jedoch zunächst im Kontext der Wirtschaftsinformatik zu spezifizieren und abzugrenzen. Der nachfolgende Abschnitt stellt die entsprechenden Überlegungen dar und liefert im Anschluss ein Vorgehensmodell für Simulationsverfahren, auf dessen Basis die weiteren Abschnitte des Kapitels strukturiert werden.

### 6.1 Strukturierung und Abgrenzung von Simulationsverfahren

Simulation wird als Begriff sowohl in den Sozial- und Wirtschaftswissenschaften als auch in der (Wirtschafts-)Informatik und dem Ingenieurwesen verwendet, wobei letztlich alle Bedeutungen des Begriffs auf Experimente mit einem Modell der Wirklichkeit zurückzuführen sind.<sup>677</sup> Das Experiment kennzeichnet sich dabei durch

- wiederholte Ausführungen unter
- kontrollierten Bedingungen mit der
- Manipulation unabhängiger Variablen und
- die Beobachtung der Auswirkungen dieser.

<sup>675</sup> Vgl. Abschnitt 4.5 und Kapitel 5.

<sup>676</sup> Vgl. Abschnitte 2.4 und 4.1–4.4.

<sup>677</sup> Vgl. Mertens (1982): 1

Dabei steht zunächst das Erfassen und Beschreiben der betrachteten Systeme im Vordergrund.<sup>678</sup> Im Kontext der WI wird Simulation häufig im Rahmen des *Operations Research* (OR)<sup>679</sup> verwendet und auch als ein Teilgebiet des OR betrachtet.<sup>680</sup> Entsprechend der Zielsetzung des OR dient die Simulation von Systemen zur Verbesserung der selbigen und nicht nur dem Verständnis dieser.<sup>681</sup> Sie kommt vorwiegend dann zum Einsatz, wenn

- vollständige, analytische Modelle der Realwelt zu komplex, zu aufwändig oder zu gefährlich zu erstellen oder zu handhaben sind,
- stochastische Elemente oder Prozesse des Systems eine bedeutende Rolle spielen,
- zeitliche Verläufe im Vordergrund stehen, deren Wiederholbarkeit unter gleichen Voraussetzungen in der Realität nicht herstellbar ist oder
- das zu untersuchende System in einem oder mehreren Aspekten ‚neu‘ ist, sodass kein Rückschluss auf ein zu erwartendes Verhalten des Systems möglich ist.<sup>682</sup>

Der Einsatz einer prototypischen Implementierung eines CroCoBIN für den beschriebenen Einsatzzweck entspricht dabei allen genannten Punkten – insbesondere der stochastischen Natur der Systemvariablen (z. B. bei der PMQ<sup>683</sup>) und der Neuartigkeit des Systems. Es lässt sich hierdurch auch die Abgrenzung zu Optimierungsmodellen herleiten, da eine vollständige Abdeckung des Problemereichs – u. a. aufgrund zu vieler notwendiger Simplifikationen – nicht möglich ist. Diese stellt jedoch ein starkes Charakteristikum analytischer, eindeutig ‚lösbarer‘ Optimierungsmodelle dar.<sup>684</sup>

<sup>678</sup> Vgl. Biethahn/Lackner/Range (2004): 5. Die Ausführungen zu Experimenten basieren auf ZIMMERMANN, der die genannten Eigenschaften als ‚Eigene Definition‘ kennzeichnet (vgl. Zimmermann (1972): 37).

<sup>679</sup> Operations Research ist ein wissenschaftlicher Ansatz, mit beschränkten Ressourcen ein gegebenes System möglichst gut zu erstellen und zu betreiben (vgl. Winston (2005): 1).

<sup>680</sup> Vgl. Domschke/Drexel (2011): 1; Winston (2005): 1145.

<sup>681</sup> GILBERT/TROITZSCH nennen noch weitere Einsatzarten der Simulation, darunter ‚Ersatz menschlicher Fähigkeiten‘, ‚Training‘, ‚Unterhaltung‘ sowie ‚Entdeckung und Formalisierung‘, die vorwiegend im sozialwissenschaftlichen Kontext untersucht werden (vgl. Gilbert/Troitzsch (2005): 4–6). Für die vorliegende Arbeit werden diese nicht weiter betrachtet.

<sup>682</sup> Vgl. Dangelmaier/Laroque (2014); Domschke/Drexel (2011): 225; Rabe (1998): 7; Winston (2005): 1145.

<sup>683</sup> Vgl. Abschnitt 5.4.

<sup>684</sup> Vgl. Domschke/Drexel (2011): 225; Winston (2005): 1145.

Aus wissenschaftstheoretischer Sicht ist die Simulation eine anerkannte Forschungsmethode der WI.<sup>685</sup> Sie lässt sich als Instrument verstehen, das dem Verständnis und der Prognose menschlichen und systemischen Verhaltens dient. Dabei reicht ein Simulationsaufbau aus, der das (erwartete) Verhalten nachbildet, das untersuchte System selbst und seine Gesetzmäßigkeiten aber nicht vollständig transparent abbildet.<sup>686</sup> BOSSEL bezeichnet dies als ‚black box‘-Ansatz, der zwangsläufig existierende Verhaltensdaten voraussetzt.<sup>687</sup> Da diese für CroCoBIN-basierte Netzwerke aus den genannten Gründen nicht existieren, muss ein – das System selbst möglichst zutreffend abbildender – ‚glass box‘-Ansatz verwendet werden. Die Basis dessen bildet immer eine strukturierte Modellentwicklung.<sup>688</sup> Abb. 52 stellt das in der vorliegenden Arbeit verwendete Vorgehensmodell zur Erstellung und Verwendung von Simulationen dar.<sup>689</sup>

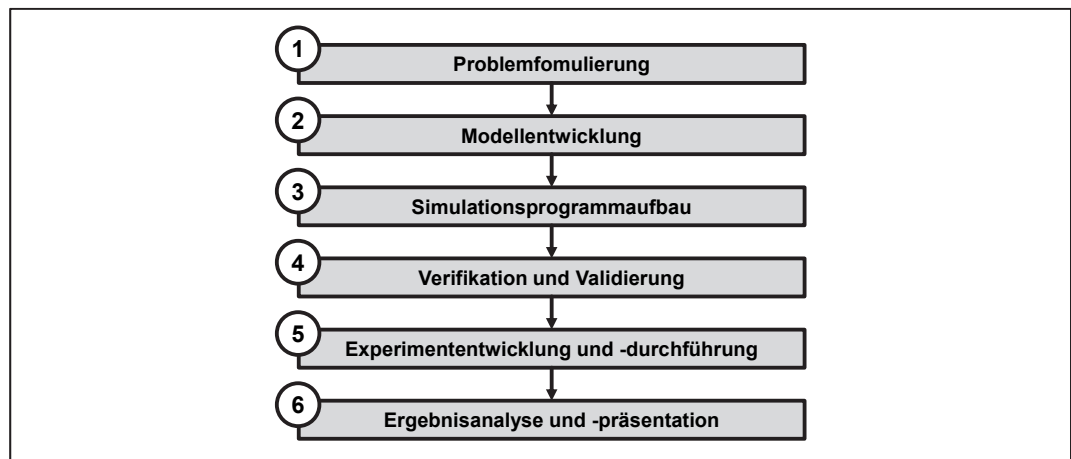


Abb. 52: Vorgehensmodell zur Erstellung einer Simulation

(Quelle: In Anlehnung an Banks (1998): 16 und Law (2007): 67)

Zuerst ist eine Problemformulierung zu entwerfen, die übergreifend bereits im einführenden Text zu diesem Kapitel gegeben wurde. Sie lässt sich weiter detaillieren, indem die Art der Simulation beschrieben wird. Ziel ist die Betrachtung der Auswirkung diverser Parameteränderungen auf die Güte des Netzes. Dabei wird

<sup>685</sup> Vgl. Abschnitt 1.2.2.

<sup>686</sup> Vgl. Weber (2004): 205f.; sowie den dort zitierten Beitrag zur Zielformulierung der WI in Nagl (1999): 422.

<sup>687</sup> Vgl. Bossel (1994): 29

<sup>688</sup> Vgl. ebd.

<sup>689</sup> Die zugrunde liegenden Vorgehensmodelle (vgl. Banks (1998): 16; Law (2007): 67) verwenden die hier vorgestellten Schritte, teilen diese jedoch partiell auf, um Rücksprünge zu ermöglichen. Auf diese wurde im Sinne einer strukturierten Bearbeitung der Schritte in den folgenden Abschnitten verzichtet, die entsprechenden granulareren Einzelschritte wurden zusammengefasst.

die zeitliche Komponente zwar in der Erstellung des vollständigen Netzes behandelt – betrachtet wird der Effekt beim Eintritt eines neuen Knotens in ein bestehendes Netz – sie steht jedoch nicht im Vordergrund, da vornehmlich die finale Konstellation eines Netzwerks mit einer gegebenen Größe, die selbst wiederum einen Parameter darstellt, bewertet wird. Es handelt sich daher um eine statische Simulation. Da das System Zufallsvariablen enthält, z. B. um unterschiedliche Prognosegüten abzubilden, handelt es sich weiterhin um eine stochastische Simulation.<sup>690</sup> Sie ist im Übrigen von einer reinen *Sensitivitätsanalyse*<sup>691</sup> abzugrenzen, die nur einen kleinen Teil der durchgeführten Experimente beschreibt, u. a. weil unterschiedliche Algorithmen zum Netzwerkaufbau eingesetzt werden.<sup>692</sup>

Im zweiten Schritt wird das Modell der Simulation entwickelt. Es basiert in der vorliegenden Arbeit auf den Überlegungen zum Fallbeispiel.<sup>693</sup> Die Beschreibung des umgesetzten Modells ist Gegenstand von Abschnitt 6.2. Die Darstellung der prototypischen Implementierung (dritter Schritt) erfolgt im anschließenden Abschnitt 6.3.

Im nachfolgenden vierten Schritt, der Prüfung, muss unterschieden werden in die Evaluation des Modells und die der Software, d. h. der Instanz.<sup>694</sup> Dabei ist wiederum zu unterscheiden nach der Verifikation, also der Bestimmung einer erwartungsgemäßen Funktion aus Sicht des Anwenders, bzw. Erstellers, und der Validierung, d. h. der Sicherstellung der Angemessenheit von Modell und Instanz für den gegebenen Einsatzzweck.<sup>695</sup>

Das Modell ist ein Artefakt, dessen Verifikation durch die dargelegte Anwendung angemessener Entwicklungsgrundsätze<sup>696</sup> und die formale und inhaltliche Nähe zu

<sup>690</sup> Vgl. zur Klassifikation Winston (2005): 1147.

<sup>691</sup> Es „[...] werden unter dem Begriff *Sensitivitätsanalyse* (SA) Verfahren bezeichnet, die Kenngrößen ermitteln, welche den Zusammenhang zwischen der Varianz der Eingangsgrößen  $x = (x_1, \dots, x_{n_f})$  und der Varianz der Ausgangsgröße  $y$  ermitteln.“ (Siebertz/van Beber/Hochkirchen: 247). Die Hervorhebungen entsprechen dem Original.

<sup>692</sup> Ebenfalls nicht zutreffend ist – trotz des Testes von Parametern – der Begriff *Parametertest*. Als Parametertests werden statistische Tests einzelner Parameter bezeichnet, die Hypothesen über deren Eigenschaften be- oder widerlegen sollen (vgl. für eine Einführung in die Thematik Kohn (2005): 375–388). Die entsprechenden Tests werden im Zusammenhang mit den Auswertungen zur (Nicht-)Normalverteilung von RQ-Werten und weiteren Simulationsergebnissen erneut thematisiert (vgl. Abschnitt 6.4.2).

<sup>693</sup> Vgl. Abschnitte 2.4 und 3.3.4.

<sup>694</sup> Vgl. Pfeiffer/Niehaves (2005): 461–463.

<sup>695</sup> Vgl. Gilbert/Troitzsch (2005): 22f.

<sup>696</sup> Vgl. Abschnitt 4.5.1.



akzeptierten Modellen gegeben ist. Seine Validierung erfolgt eben durch die Simulation, d. h. die Instanziierung. Es werden dazu *Modellvergleiche*<sup>697</sup> vorgenommen, indem wesentliche Komponenten, Parameter und Vorgehensweisen (stark) modifiziert und in Teilen vollständig ausgetauscht werden, sodass im Rahmen des CroCoBIN-Referenzmodells mehrere Modellinstanzen entstehen.<sup>698</sup> Zudem können naive Annahmen<sup>699</sup> genutzt werden, um *Plausibilitätstests* und allgemeine *a-posteriori-Beobachtungen* durchzuführen.<sup>700</sup> Es ist dazu ergänzend festzustellen, dass Modelle ihrer Natur nach immer eine vereinfachte Form der Realität abbilden. Unabhängig von der Häufigkeit der Übereinstimmung von Aussagen, die über das Modell getroffen werden können, mit Aussagen, die über das reale Vorbild getroffen werden können, ist daher eine vollständige Sicherheit über die semantische Korrektheit, d. h. eine in Struktur und Verhalten zutreffende Abbildung der Realität im Modell, nie erreichbar.<sup>701</sup>

In Bezug auf die prototypische Implementierung sind sowohl Verifikation als auch Validierung einfacher oder zumindest eindeutiger durchführbar. Für die Software-Entwicklung existieren anerkannte Vorgehensweisen zur Sicherstellung einer ordnungsgemäßen Funktion, die im Rahmen der Beschreibung des Simulationsprogrammaufbaus diskutiert werden.<sup>702</sup> Eine Validierung kann hier z. B. durch Extremwerttests und grafische Abbildungen des verwendeten Modells vorgenommen werden.<sup>703</sup>

Die eigentliche Durchführung der Simulation beinhaltet die Wahl der zu verwendenden Parameterbereiche, die Identifikation geeigneter Algorithmen und Verknüpfungsnormen sowie die Analyse unterschiedlicher Netzwerktypen und -größen. Abschnitt 6.4 bietet eine ausführliche Darstellung, Analyse und Interpretation der so erzielten Simulationsergebnisse und bildet damit die letzten beiden Schritte des Vorgehensmodells der Simulation ab.

---

<sup>697</sup> Vgl. Bungartz et al. (2009): 16.

<sup>698</sup> Die entsprechenden Konfigurationsmöglichkeiten und verwendeten Konfigurationen, z. B. in Bezug auf Nachbarschaftswahlen, sind Gegenstand der Abschnitte 6.2–6.4.

<sup>699</sup> Eine naive Annahme ist z. B.: ‚Je dichter das Netzwerk, desto besser die NQ.‘ (vgl. Abschnitt 5.4).

<sup>700</sup> Vgl. Bungartz et al. (2009): 16.

<sup>701</sup> Vgl. Holl/Auerachs (2004): 383f.

<sup>702</sup> Vgl. Abschnitt 6.3.1.

<sup>703</sup> Vgl. Banks (1998): 22–24. Die entsprechenden Umsetzungen in der Implementierung werden in den Abschnitten 6.3.3 und 6.4 beschrieben.

## 6.2 Modell des Fallbeispiels

Das für die Simulation zu verwendende Modell basiert auf den fachlichen Annahmen des vorgestellten Fallbeispiels.<sup>704</sup> Es stellt ein Netzwerk kleiner und mittelgroßer Unternehmen dar, die im Bereich F&E miteinander kooperieren wollen. Dazu tauschen sie Informationen über abgeschlossene Entwicklungsprojekte miteinander aus. Das entstehende P2P-BI-Netzwerk folgt den Grundlagen des CroCoBIN-Referenzmodells. Für die Betrachtung der in den vorangegangenen Kapiteln und Abschnitten diskutierten Größen wird ein hinreichender Ausschnitt des Referenzmodells implementiert. Abb. 53 stellt den Umfang des Ausschnitts dar, wobei alle schwarz hinterlegten Bestandteile berücksichtigt werden. Nicht aufgenommen werden u. a. die interne Datenhaltung, da diese für das Netzwerk nicht zugänglich ist,<sup>705</sup> und die Caching-Verfahren, da diese in erster Linie zu Performance-Verbesserungen dienen, jedoch keine Aussagen über die Grundstruktur des Netzes treffen.<sup>706</sup>

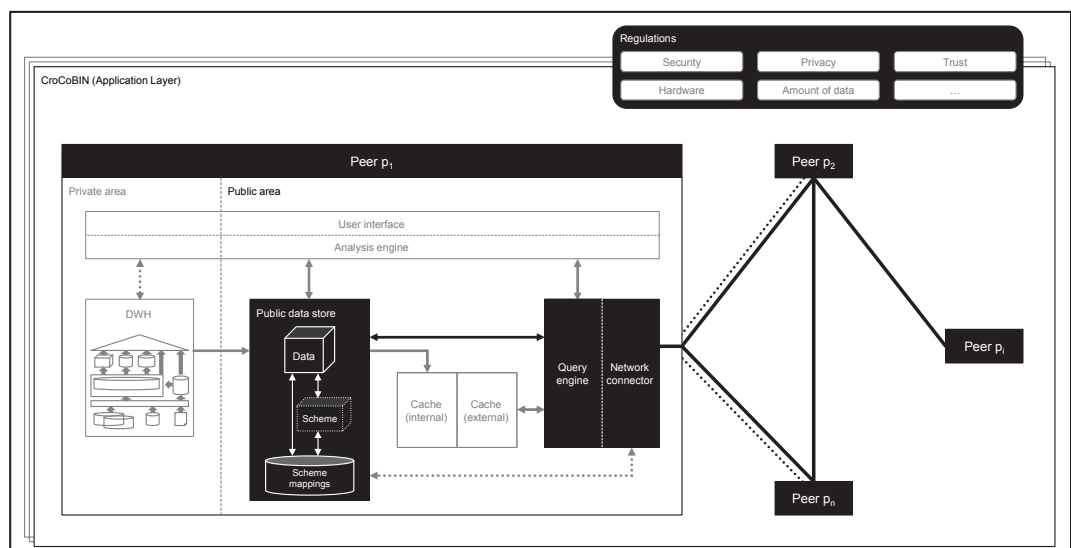


Abb. 53: Umfang des Modells bezogen auf das CroCoBIN-Referenzmodell  
(Quelle: Eigene Darstellung)

<sup>704</sup> Vgl. Abschnitt 2.4.

<sup>705</sup> Vgl. Abschnitt 4.5.2.2.

<sup>706</sup> Vgl. Abschnitt 4.2.2. Das Einbeziehen von vorgehaltenen Daten kann zwar im Zeitverlauf die ‚Datentransportwege‘ verändern – insbesondere verkürzen –, dies ist jedoch nur möglich, wenn ein existierendes, funktionierendes Netz bereits erfolgreich Daten übertragen hat, sodass hier keine Erkenntnisse über grundlegende Strukturen, sondern über Verbesserungen bereits implementierter Netze zu erlangen sind.

Das Modell bildet ein P2P-Netzwerk nach, wobei aufgrund des Fokus auf die Anwendungsschicht<sup>707</sup> keine tatsächliche Implementierung eines P2P-Netzwerks erforderlich ist, sondern eine lokale Abbildung eines solchen mit einer beliebigen Menge an Teilnehmern verwendet werden kann, solange die Eigenschaft unabhängiger Peers aufrecht erhalten wird. Typische Netzgrößen liegen dabei zwischen wenigen dutzend bis hundert Teilnehmern vor.<sup>708</sup> Jeder Peer hält einen öffentlich zugänglichen, multidimensional strukturierten Datenbestand vor. Zur besseren Vergleichbarkeit wird im Modell ein übergreifendes Schema verwendet, das durch randomisierte Ausprägungen, z. B. einen Wechsel zwischen alternativen Hierarchien, messbar abweichende Peer-Schemata ermöglicht. Abb. 54 zeigt das verwendete Master-Schema des Modell-Würfels.<sup>709</sup> Die tatsächliche Umsetzung peer-abhängiger Schemata wird im Zuge der Implementierungsbeschreibung erläutert.<sup>710</sup>

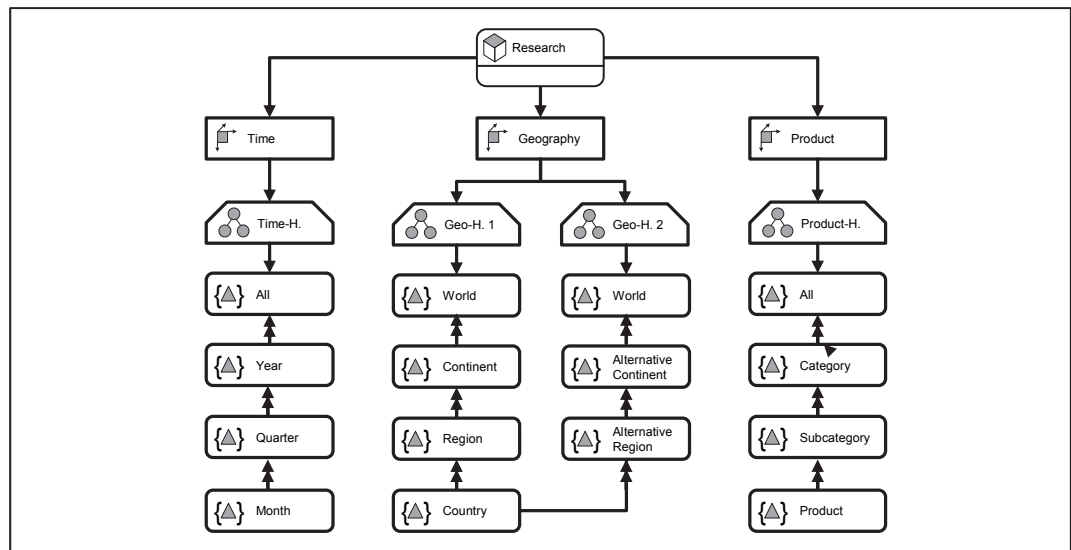


Abb. 54: ADAPT-Darstellung des Würfels zum Fallbeispiel

(Quelle: Eigene Darstellung)

<sup>707</sup> Vgl. Abschnitt 4.5.2.1.

<sup>708</sup> Vgl. Abschnitt 2.4; sowie den einleitenden Text zu Kapitel 5.

<sup>709</sup> Für die Beschreibung der Ebenen und Elemente des Modells wird – wie bereits in anderen Modellen und der gesamten Implementierung – auf englischsprachige Begriffe zurückgegriffen. Insbesondere in der Implementierung ist dies durch die starke Stellung des Englischen in der Software-Entwicklung zu begründen: Die 20 meistverwendeten Programmiersprachen der Welt stehen für ca. 75% des erschaffenen Quellcodes und basieren alle auf englischsprachigen Schlüsselwörtern (vgl. TIOBE (2015)).

<sup>710</sup> Vgl. dazu und zu Überlegungen der Datenmodellkomplexität Abschnitt 6.3.2.1. Die Auswahl von drei Dimensionen bietet eine Größe, die zum einen eine ausreichende Variabilität in der Schema-Generierung zulässt, zum anderen noch mit vertretbarem Aufwand manuell erfassbar und überprüfbar ist, was Verifikation und Validierung von Modell und Instanz ermöglicht (vgl. dazu Abschnitt 6.1).

Abschließend sind allgemein die variierbaren Parameter der Verbindungen sowie die Rahmenbedingungen des Netzes festzulegen. Basierend auf den Überlegungen der Abschnitte 4.5 und 5.2 bis 5.4 zählen dazu:

- Die grundlegende Netztopologie,<sup>711</sup>
- die Größe des Netzwerks,
- eine Zielgröße der Verbindungszahl,
- die Strategie der Nachbarschaftswahl sowie die dabei verwendete Prognosegüte für die Verbindungen,
- die Auswahl der Parameter der MQ und
- die Auswahl und Konfiguration einer geeigneten Kombinationsfunktion.

Die im Einzelnen auswählbaren Konfigurationen werden im Rahmen der implementierten Lösung,<sup>712</sup> die tatsächlich verwendeten Konfigurationen im Rahmen der Beschreibung der Simulationsdurchführung vorgestellt und diskutiert.<sup>713</sup>

### **6.3 CroCoSIM: Prototypische Implementierung einer Simulation für Peer-to-peer-Business-Intelligence-Netzwerke**

Die Simulation von CroCoBIN-Netzwerken ist auf Basis bestehender Simulationssoftware nicht – oder nicht mit angemessenem Aufwand – unter Berücksichtigung aller genannten Modelleigenschaften und -parameter möglich. Zwar existieren Programme, die umfangreiche Techniken zur Untersuchung von Graphen beinhalten, diese bieten jedoch keine parametrisierbaren Mehrfachsimulationen an, sondern fokussieren zumeist unterschiedliche Graphendarstellungen zur visuellen Analyse.<sup>714</sup> Typischerweise im Bereich OR eingesetzte Simulationssoftware hingegen dient eher der Betrachtung von Geschäftsprozessen als Netzdarstellungen.<sup>715</sup> Netzspezifische Simulationen wiederum berücksichtigen zwar Transportprotokolle und Netzwerktopologien, nicht aber den Aufbau und die Funktionen –

---

<sup>711</sup> Dies beinhaltet insbesondere die Wahl von Zufallsnetzwerken oder skalenfreien Netzwerken (vgl. Abschnitt 5.2.2).

<sup>712</sup> Vgl. Abschnitt 6.3.2.

<sup>713</sup> Vgl. Abschnitt 6.4.

<sup>714</sup> Als bekanntester Vertreter in der wissenschaftlichen Gemeinschaft kann *Gephi* (<http://gephi.github.io/>) gelten, das aufgrund seiner Open-Source-Eigenschaft vielfach eingesetzt wird (vgl. Heymann (2014): 614f.) Auch das aus biologischen Forschungsprojekten hervorgegangene Cytoscape (<http://www.cytoscape.org/>) ist auf Visualisierungen fokussiert.

<sup>715</sup> Vgl. OR/MS Today (2013).

und damit auch keine multidimensionale Datenhaltung – der Peers.<sup>716</sup> Zur Abbildung der umfangreichen problemspezifischen Eigenarten des Modells wurde daher mit *CroCoSIM*<sup>717</sup> eine prototypische Implementierung des Modells mit Simulationsfähigkeiten erstellt und verwendet. Die folgenden Abschnitte beschreiben die grundlegende Systemarchitektur, die Konfigurationsmöglichkeiten, die Besonderheiten bei der Abbildung der Schemaüberführung,<sup>718</sup> die Sicherstellung der Verwendung korrekter Netzwerktopologien und die existierenden Auswertungsmöglichkeiten.

### 6.3.1 Systemarchitektur und Implementierung

CroCoSIM wurde vollständig auf Basis der Programmiersprache C# und des Microsoft®-.NET-Frameworks 4.5 entwickelt, was aufgrund der weiten Verbreitung der Sprache eine hohe Überprüfbarkeit und Wiederverwendbarkeit sichert.<sup>719</sup> Die Architektur der Software orientiert sich dabei an dem in der Softwareentwicklung etablierten *Model-View-Controller*(MVC)-Muster.<sup>720</sup> Abb. 55 zeigt ein Klassendiagramm der Software, das die Zuordnung der erstellten Komponenten zu den Bereichen des MVC-Musters darstellt.<sup>721</sup> Eine vollständige Dokumentation der Software und der verwendeten Softwarekomponenten Dritter bietet Anhang 1.

<sup>716</sup> Vgl. Siraj/Gupta/Badjugar (2012): 201f.

<sup>717</sup> CroCoSIM wird als zu CroCoBIN analoge Bezeichnung der entsprechenden Simulationskomponente gewählt. Vgl. für die begriffliche Einordnung die Hinweise zu CroCoBIN in Fußnote 486.

<sup>718</sup> Vgl. zu den Besonderheiten und der Wichtigkeit von angemessenen Schemaüberführungen Abschnitt 4.3.

<sup>719</sup> Vgl. zu der Verbreitung TIOBE (2015).

<sup>720</sup> Das MVC-Muster sieht eine Trennung der Komponenten einer Software in drei übergeordnete Bereiche vor. Die Komponenten des Modells stellen das durch die Software abzubildende (Daten-)Modell dar, die Komponenten des Controllers übernehmen Programmlogik und -steuerung, die Komponenten der View dienen der Präsentation des Modells sowie der direkten Interaktion mit dem Nutzer (vgl. Ludwig/Lichter (2010): 433–435; Sommerville (2012): 192f.). Für eine ausführliche Darstellung der verwendeten Begrifflichkeiten (‘Framework’, ‘Architektur’, etc.) vgl. die Hinweise zur Entwicklung des Referenzmodells in Abschnitt 4.5.1.

<sup>721</sup> Klassendiagramme stellen ein im Kontext der objektorientierten Programmierung übliches Dokumentationswerkzeug dar, um die Objekte und ihre Beziehungen zu verdeutlichen. Sie werden üblicherweise mit Hilfe der *Unified Modeling Language* (UML) erstellt, die auch in der vorliegenden Arbeit verwendet wird (zu UML und Klassendiagrammen vgl. Object Management Group (2010): 95–99). Das hier vorliegende Klassendiagramm kann als ‚vereinfacht‘ bezeichnet werden, da es lediglich die allgemeinen Vererbungen und Aggregationen, darstellt, jedoch keine weitergehenden Informationen wie z. B. Multiplizitäten der Verbindungen beinhaltet.

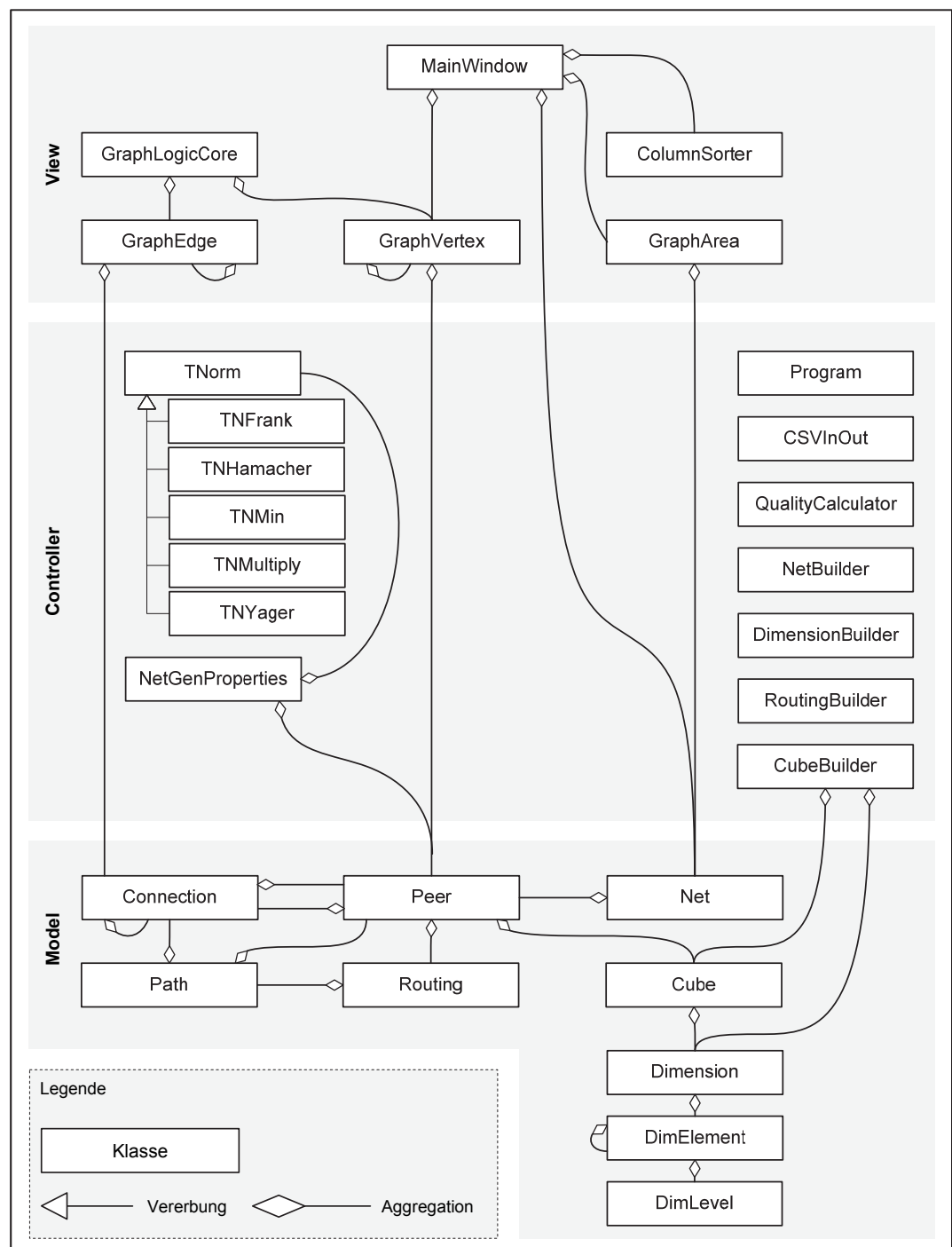


Abb. 55: Vereinfachtes UML-Klassendiagramm von CroCoSIM

(Quelle: Eigene Darstellung)

Zur Evaluation der Software wurden die für Instanziierungen üblichen Methoden verwendet.<sup>722</sup> Ein durchgehendes *Code-Review*, d. h. eine Überprüfung des erstellten Quelltextes, ermöglichte die Beseitigung eventueller Fehler bereits wäh-

<sup>722</sup> Für die entsprechende Aufstellung vgl. Pfeiffer/Niehaves (2005): 463. Die Autoren weisen hier die ‚Verifikation‘ als einzelnen Bestandteil aus, sie wird im Kontext der Software-Entwicklung jedoch auch als Oberbegriff verstanden und beinhaltet z. B. die hier beschriebenen Testverfahren (vgl. Halang/Konakovsky (2013): 133f.).

rend des Entwicklungsprozesses. Ein solches Review ist jedoch selbst bei der mehrfachen Durchführung unabhängiger Entwickler nicht ausreichend zur Sicherstellung einer korrekten Funktion.<sup>723</sup> Daher wurden sowohl dynamische *Testverfahren* als auch Methoden der statischen *Code-Analyse* eingesetzt. Der Test stellt dabei die am häufigsten eingesetzte Methode zur Verifikation dar und lässt sich in Unit-Tests, Integrationstests, Systemtests und Abnahmetests unterscheiden.<sup>724</sup> Um eine umfangreich abgesicherte Funktionsfähigkeit der Software nachzuweisen, wurden sowohl *White-Box-Tests* – zur Überprüfung der korrekten Struktur – als auch *Black-Box-Tests* – zur Überprüfung der korrekten Funktion – durchgeführt.<sup>725,726</sup> CroCoSIM weist in dem für die durchgeführten Simulationen verwendeten Entwicklungsstand keinerlei mit diesen Testmethoden nachgewiesene Fehler auf. Ergänzend wurde eine automatische, unabhängige Code-Analyse durchgeführt, mit der nachgewiesen werden konnte, dass keine Probleme im Programm vorliegen.<sup>727</sup> Die korrekte Abbildung der Funktionen entspricht dabei bereits in Teilen einer Validierung der Software, da CroCoSIM eine für einen speziellen Einsatzbereich konstruierte Software ist, deren korrekte Funktion gleichzeitig eine dem Anwendungsfall angemessene Abbildung des Problembereichs belegt.<sup>728</sup>

<sup>723</sup> Vgl. Edmundson et al. (2013): 203–210.

<sup>724</sup> Vgl. Hoffmann (2013): 157–159. Für die vorliegende Arbeit sind nur die beiden zuerst genannten Prüfebene relevant, da eine Absicht zur Veröffentlichung der Software und eine damit verbundene Notwendigkeit zur Funktionalitätsüberprüfung mit einer Abnahmezielsetzung nicht vorliegen.

<sup>725</sup> Verwendet wurden z. B. Tests der Bedingungsüberdeckung (White-Box-Test), um für die verwendeten Metriken (MQ, RQ, etc.) eine vollumfängliche Erreichbarkeit des Zielwertebereichs sicherzustellen. Nach der Sicherstellung dieser ‚möglichen‘ Erreichbarkeit wurde mit Hilfe von Grenzwertbetrachtungen (Black-Box-Test) die ‚tatsächliche‘ Erreichung der Werte bei entsprechenden Parametern sichergestellt (vgl. zu den Testmethoden und ihrer Einordnung Hoffmann (2013): 180–183, 214–216).

<sup>726</sup> Die Unterscheidung der genannten Testarten (White-Box-Test, Black-Box-Test) ist üblich, wird jedoch vor dem Hintergrund immer umfangreicherer Programme und Testmethoden diskutiert (vgl. Liggesmeyer (2009): 40). Für die vorliegende Arbeit wird sie aufgrund der prototypischen Art der Software als genügend angenommen.

<sup>727</sup> Verwendet wurde die umfangreiche Code-Analysefunktion der Entwicklungsumgebung Microsoft® Visual Studio (zu den entsprechenden Funktionen vgl. Kamel (2013); Microsoft (2013)). Der Einsatz entsprechender Werkzeuge ist in der Software-Entwicklung gleichermaßen üblich wie wichtig, da so eine objektive Überprüfung des Programms gewährleistet wird (vgl. Hoffmann (2013): 247–249).

<sup>728</sup> Vgl. Abschnitt 6.1. Die durchgeführten Grenzwertbetrachtungen/Extremwerttests stellen zudem nach BANKS Methoden der Validierung dar (vgl. Banks (1998): 23f.).

### 6.3.2 Umfang und Konfigurationsmöglichkeiten

Basierend auf den Überlegungen zum Modell des Fallbeispiels,<sup>729</sup> stellt CroCoSIM sowohl die Möglichkeit einzelner als auch mehrfacher Netzsimulationen bereit.<sup>730</sup> Dabei ist insbesondere für die Batch-Verarbeitung ein definierbarer, veränderlicher Seed<sup>731</sup> berücksichtigt, der zur Validierung der Robustheit der Ergebnisse grundsätzlich notwendig ist.<sup>732</sup> Die folgenden Abschnitte stellen die für die einzelnen Varianten und Bestandteile der Simulation verfügbaren Konfigurationsmöglichkeiten dar.

#### 6.3.2.1 Abbildung der Schemaüberführung

Zur Entwicklung einer realitätsnahen MQ-Simulation wird für jeden Peer ein Modell-entsprechender Datenwürfel erzeugt. Dazu wird auf einen Master-Würfel zurückgegriffen, der randomisiert verkleinert wird, sodass zum einen eine Vielzahl unterschiedlicher Würfel mit (deutlich) abweichenden Granularitäten in den Dimensionen erzeugt wird,<sup>733</sup> zum anderen aber durch die Verwendung identisch bezeichneter Elemente ein vollautomatischer Matching-Prozess ermöglicht wird, bzw. ein Mapping bereits durch die Definition der Schemata gegeben ist. Im Simulationsprozess steuern drei Parameter die Reduktion des Master-Schemas, indem Wahrscheinlichkeiten für die Auslassung von Elementen oder vollständigen Ebenen sowie zur Verwendung alternativer Hierarchien bei der Bildung der Würfel festgelegt werden. Tab. 13 stellt die Parameter und ihre Effekte einander gegenüber.

<sup>729</sup> Vgl. Abschnitt 6.2.

<sup>730</sup> Das mehrfache Verarbeiten mit unterschiedlichen Parameterkonfigurationen wird, in Anlehnung an den gängigen englischen Begriff, im Weiteren als *Batch*(-Verarbeitung) bezeichnet. Die einzelnen Varianten der Batch-Verarbeitung sind in Anhang 1 dokumentiert und basieren vollständig auf der für Einzelsimulationen verwendeten algorithmischen Basis, sodass jedes Ergebnis einer Mehrfachsimulation auch einzeln geprüft werden kann und keine fehlerhaften Abweichungen durch Verwendung unterschiedlicher Funktionen entstehen können.

<sup>731</sup> Zur Abbildung der stochastischen Simulationselemente werden (Pseudo-)Zufallszahlgeneratoren verwendet, die einen gegebenen Wert (den *Seed*) als Startwert zur Generierung einer Reihe von (Pseudo-)Zufallszahlen verwenden. Unterschiedliche Seeds führen dabei zu abweichenden Zahlenreihen (vgl. Goldreich/Goldwasser/Micali (1986): 796). Die mehrfache Verwendung desselben Seeds führt dabei zu identischen Reihen, sodass eine Nachvollziehbarkeit und Vergleichbarkeit unterschiedlicher Simulationen erzielt werden kann.

<sup>732</sup> Vgl. Gilles (2007): 267.

<sup>733</sup> Zur Erstellung des Masterwürfels lässt sich eine Microsoft® Excel-Datei verwenden (vgl. Anhang 1), die für jede zu verwendende Dimension die Elemente und Ebenen vorhält und somit – bei Bedarf – die Verwendung unterschiedlicher Würfelschemata erlaubt.



Parameter	Beschreibung
<b>P(skip_elem)</b>	Definiert die Wahrscheinlichkeit, dass ein bestimmtes Element des Master-Würfels nicht mit in den zu erstellenden Würfel des Peers aufgenommen wird. $P(skip\_elem) \in [0; 1]$
<b>P(skip_level)</b>	Definiert die Wahrscheinlichkeit, dass eine bestimmte Ebene des Master-Würfels nicht mit in den zu erstellenden Würfel des Peers aufgenommen wird. $P(skip\_level) \in [0; 1]$
<b>P(2nd_hier)</b>	Definiert die Wahrscheinlichkeit, dass – sofern existent – die alternative Hierarchie einer Dimension des Master-Würfels für den zu erstellenden Würfel des Peers verwendet wird. $P(2nd\_hier) \in [0; 1]$

Tab. 13: Parameter zur Definition der Schemagenerierung

(Quelle: Eigene Darstellung)

Das implizit geschaffene Mapping berücksichtigt somit nur eindeutig überführbare Elemente und keine Splits oder Umwidmungen.<sup>734</sup> Es erlaubt aber die Betrachtung heterogener Hierarchien, was zum einen in automatischen Matching-Verfahren nicht immer gewährleistet werden kann, zum anderen durch die daraus resultierende Änderung von Elternelementen die Konsistenzprüfung als Teil der MQ-Bildung in der Simulation ermöglicht.<sup>735</sup> Die Abbildung komplizierterer Überführungen (z. B. durch Splits) bietet für die vorliegende Simulation keinen Mehrwert, da sie zwar die Erstellung des Mappings, nicht aber den Werteraum der MQ beeinflusst. Durch die physische Implementierung einer multidimensionalen Datenstruktur ist bereits eine beliebige Errechnung der MQ und der Teilqualitäten umsetzbar, die die durch Hierarchiebildung und Multidimensionalität entstehende Komplexität berücksichtigt. Eine weitere Komplexitätserhöhung würde somit nicht den Erkenntnisgewinn erhöhen, sondern nur die zugrunde liegenden Werte verändern.<sup>736</sup>

<sup>734</sup> Vgl. zu den beschriebenen Besonderheiten Abschnitt 4.3.4.

<sup>735</sup> Vgl. Abschnitte 4.3.3 und 5.3.1.

<sup>736</sup> Bei der Generierung der Dimensionen und Würfel wird zudem immer vorausgesetzt, dass das oberste Element jeder Dimension erhalten bleibt, damit in jedem Fall eine minimale Überführbarkeit gegeben ist. Im gegenteiligen Fall läge überhaupt kein Mapping vor. Damit bestünde aber auch keine semantische Überführung der Peers, was per Definition eines CroCoBIN ausgeschlossen ist (vgl. Abschnitt 4.5.2.1). Die minimale TMQ beträgt damit im Simulationsmodell nur näherungsweise 0 (da immer eine minimale Übereinstimmung vorliegt) und liegt bei  $\frac{1}{\prod_i |D_i|} = \frac{1}{69 \times 87 \times 30} \approx 0,00000553$ , was auch die theoretische Untergrenze der MQ darstellt und die Aussagekraft im Wertebereich  $[0;1]$  nicht relevant einschränkt.

### 6.3.2.2 Bestimmung von Netzaufbau und Nachbarschaftswahl

Die Generierung von Netzen geschieht in CroCoSIM auf Basis zweier Zieltopologien. Das erstellte Netz genügt dabei entweder der Struktur eines Zufallsnetzwerks oder eines skalenfreien Netzwerks.<sup>737</sup> In beiden Fällen wird das Netzwerk iterativ vergrößert, d. h. es wird ein Startknoten oder Startnetz erzeugt, zu dem weitere Peers hinzugefügt werden, bis die endgültige Größe des Netzwerks erreicht wird. Jeder hinzugefügte Knoten trifft dabei die Auswahl seiner Nachbarschaft basierend auf den Auswahlparametern und unter Berücksichtigung der Zieltopologie. Im Folgenden werden beide Ansätze vorgestellt, vorab stellt Tab. 14 die Parameter zum Netzaufbau vor.

Parameter	Beschreibung
<b>Basic algorithm</b>	Erstellt entweder ein Zufallsnetzwerk ( <i>Random</i> ) oder ein skalenfreies Netzwerk ( <i>Scale-free</i> ), basierend auf dem Preferential-Attachment-Algorithmus von BARABASI/ALBERT („BA-Algorithmus“). <sup>738</sup>
<b>Net size</b>	Definiert die Netzgröße des vollständig entwickelten Netzes, dessen Eigenschaften ausgewertet werden. $net\_size \in \mathbb{N}^{>0}$
<b>Target degree</b>	Nur für Zufallsnetzwerke: Definiert den angestrebten Grad jedes Knotens im Netzwerk und damit für eine vorgegebene Netzgröße die Dichte des vollständigen Netzwerks. $target\_degree \in \mathbb{N}^{>1}$
<b>m</b>	Nur für skalenfreie Netzwerke: Definiert die Anzahl an Knoten, mit denen sich jeder neu hinzugefügte Knoten verbindet und dadurch auch die Größe $m_0$ eines initialen Kernnetzes. $m \in \mathbb{N}^{>0}, m_0 = m + 1$ Es ergibt sich dadurch für eine vorgegebene Netzgröße die Dichte des vollständigen Netzwerks $G$ , wobei zur Vergleichbarkeit mit einem Zufallsnetzwerk $H$ gilt: $d(G_{2m}) \cong d(H_{target\_degree})$ .
<b>Core net size</b>	Nur für Zufallsnetzwerke: Definiert die Größe des vollständig verbundenen, initialen Kernnetzwerks. $corenet\_size \in \mathbb{N}^{>0}$
<b>Seed</b>	Definiert den Seed, mit dem der Zufallsgenerator zur Erzeugung der Netze initialisiert wird. Seeds weiterer Zufallsgeneratoren (z. B. zur Erzeugung der Peer-Würfel) sind funktional von diesem abhängig.

Tab. 14: Parameter zur Gestaltung des Netzaufbaus

(Quelle: Eigene Darstellung)

Die Definition eines initialen, vollständig verbundenen Kernnetzes ist zur Erstellung eines Zufallsnetzwerks nicht zwingend notwendig, sie bildet aber die Mög-

<sup>737</sup> Vgl. zu den Netzwerktopologien Abschnitt 5.2.2.

<sup>738</sup> Vgl. zu dem Vorgehen Albert/Barabási (2002): 71. Das Grundprinzip basiert auf der Auswahl von Zielknoten mit steigender Wahrscheinlichkeit in Abhängigkeit von deren Gradanzahl.

lichkeit – analog zum Parameter  $m$  in skalenfreien Netzwerken – eine Kerngruppe an Peers zu definieren, die ein Netzwerk begründet.<sup>739</sup>

Netzgenerierungen durch die Basisalgorithmen<sup>740</sup> führen – da sie dort nicht definiert wird – zu keiner Berücksichtigung der (prognostizierten) MQ. Da die Auswahl der Topologie jedoch nur eine Rahmenbedingung setzen soll, wurden die Basisalgorithmen um eine MQ-fokussierte Nachbarschaftswahl ergänzt. Die entsprechenden Modifikationen lassen sich dabei auf das Prinzip einer Veränderung der initialen Attraktivität eines Knotens zurückführen, was die ursprünglichen Netzeigenschaften erhält.<sup>741</sup> Dabei kann die Veränderung der Attraktivität auch dynamisch erfolgen. CHEN/SHI fügen dazu eine zusätzliche Variable in den originären Algorithmus ein.<sup>742</sup> Wird die Wirkungskraft der Variable zusätzlich über einen Koeffizienten gesteuert, liegt ein ‚Two-Level-Model‘ vor, wobei ein niedriger Koeffizientenwert üblicherweise die Wirkung der Veränderungen verringert, ein hoher Koeffizientenwert sie entsprechend erhöht.<sup>743</sup> Im vorliegenden Modell wird die Attraktivität eines Peers daher nicht nur von seiner Gradzahl, sondern auch von der PMQ zwischen dem neu eintretenden Peer und dem betrachteten potenziellen Partner bestimmt.

Der in CroCoSIM verwendete modifizierte Zufallsnetzwerk-Algorithmus („modified random algorithm“ (MRA)) sowie der modifizierte BA-Algorithmus verwenden unterschiedliche Parameter zur Steuerung. Tab. 15 zeigt alle Parameter mit einer Zuordnung zu den Algorithmen und einer Wirkungsbeschreibung. Eine Be-

<sup>739</sup> Die Verwendung der *corenet\_size* lässt sich im Rahmen der Betrachtung von Zufallsnetzwerken auf Überlegungen zu unterschiedlich „großen“ Seeds zurückführen, die in anderen Arbeiten z. B. bei der Entfernung oder Aktivierung von Knoten in Netzen verwendet werden (vgl. Gleeson/Cahalane (2007): 1). Im Allgemeinen bildet sie die vorgestellten Überlegungen zu der notwendigen „kritischen Masse“ bei einem initialen Netzwerkaufbau ab (vgl. Abschnitte 1.1 und 6.4.3.2).

<sup>740</sup> Während für skalenfreie Netzwerke mit dem BA-Algorithmus ein Vorgehen definiert wird, das iterativ Knoten hinzufügt, werden Zufallsnetzwerke üblicherweise mit einem kantenfreien Graphen erzeugt, dem zufällig mit einer konstanten Wahrscheinlichkeit Kanten hinzugefügt werden (vgl. Nobari et al.: 332). Dies verhindert jedoch einen fixierbaren durchschnittlichen Grad der Knoten, was die Vergleichbarkeit der Netzwerke einschränkt. Das hier verwendete Vorgehen beschränkt zwangsläufig die Generierung von Kanten auf bereits im Netz vorhandene Knoten – und verwendet somit keine konstante Wahrscheinlichkeit –, behebt diese Ungleichheit jedoch mit einer vorherigen Randomisierung der Knotenmenge (vgl. Anhang 2).

<sup>741</sup> Für skalenfreie Netzwerke ist diese Annahme bereits nachgewiesen (vgl. Albert/Barabási (2002): 77f.). Für Zufallsnetzwerke wurde sie im Rahmen der vorliegenden Arbeit überprüft und bestätigt (vgl. Anhang 4).

<sup>742</sup> Vgl. Chen/Shi (2004): 243–246.

<sup>743</sup> Vgl. Dangelchev (2004): 660f.

schreibung der Algorithmen in Pseudocode<sup>744</sup> ist in Anhang 2 und Anhang 3 dargestellt.<sup>745</sup>

Parameter	MRA	BA	Beschreibung
<b>PMQ goodness (PMQG)</b> <sup>746</sup>	X	X	Definiert die Genauigkeit, mit der bei der Überprüfung aller möglichen Paarungen die tatsächlich erreichte MQ – nicht aber die TMQ – bestimmt werden kann. Ein Wert von 0 entspricht Zufallswerten, ein Wert von 1 der exakten Vorhersage. $PMQG \in [0; 1]$
<b>Target degree spread</b>	X		Definiert die Höhe der Abweichung eines Knotengrades vom Zielwert. Ein Wert von 0 erzwingt eine vollständige Einhaltung des Zielwertes für alle Knoten im Netz. $target\_degree\_spread \in [0; 1]$
<b>Sorted node list</b>		X	Prüft die potenziellen Partnerknoten im Netz nicht nach zufälliger, sondern nach geordneter Reihenfolge und erhöht bei gleichbleibenden Einzelwahrscheinlichkeiten für jede einzelne Überprüfung die Wahrscheinlichkeit, einen Knoten mit hoher PMQ zu wählen. $sorted \in \{true; false\}$
<b><math>\mu</math></b>		X	Bestimmt den Effekt der PMQ auf die Attraktivität der im Netz bestehenden Knoten. Ein Wert von 0 ignoriert die PMQ, ein Wert von 1 setzt ihn mit der Attraktivität der höchsten Gradzahl ( $\max(d)$ ) gleich. $\mu \in \mathbb{R}^{\geq 0}$ , Gesamtattraktivität $\sim \frac{\mu \times PMQ \times \max(d) + d}{1 + \mu}$
<b>Selfish approach</b>	X	X	Kehrt, falls auf <i>falsch</i> ( <i>false</i> ) gesetzt, die Betrachtung der Qualitäten um und prüft nicht die Qualität des Datenempfangs, sondern des Datensendens. <sup>747</sup> $selfish \in \{true; false\}$
<b>MQ (<math>\alpha, \beta, \gamma</math>)</b>	X	X	Definiert den Einfluss von $MQ_s(\alpha)$ , $MQ_c(\beta)$ und $MQ_b(\gamma)$ auf die MQ (vgl. Definition (26)).

Tab. 15: Parameter zur Beeinflussung der Nachbarschaftswahl

(Quelle: Eigene Darstellung)

<sup>744</sup> Pseudocode ist eine zwischen menschlicher und maschinenlesbarer Sprache eingeordnete Form der schriftlichen Aufbereitung von Computerprogrammen, die eine natürlich-sprachliche Beschreibung des Vergehens eines Programms darstellt und um feststehende Begriffe ergänzt werden kann (vgl. Roy (2006): 2).

<sup>745</sup> Für eine vollständige Beschreibung der Algorithmen vgl. die Dokumentation in Anhang 1.

<sup>746</sup> Entsprechend den Ausführungen zur Definition der Messgrößen (vgl. Abschnitt 5.3) führt das Abkürzungsverzeichnis den deutschen Begriff ‚Prognostizierte-Mapping-Qualität-Güte‘.

<sup>747</sup> Wird ein Knoten  $p$  dem Netz  $G$  hinzugefügt, so wird an Stelle von  $MQ(p, q)$ , mit  $q \in G$ , immer  $MQ(q, p)$  verwendet. Diese Eigenschaft gilt *nur* für die Nachbarschaftswahl, nicht für die Erstellung des Routings in einem vollständig generierten Netz, da dieses gerichtete Wege interpretiert und die bereits bestehenden Verbindungen nutzt. Der Parameter bildet damit die (Nicht-)Anwendung des in Abschnitt 5.4 vorgestellten ‚gierigen‘ oder ‚selbstsüchtigen‘ Verhaltens bei der Nachbarschaftswahl ab.

Nach Aufbau des vollständigen Netzwerks wird auf Basis der Nachbarschaften und vorliegenden MQ ein Routing für jeden Peer erstellt. Das dabei verwendete Vorgehen basiert auf dem Algorithmus von Dijkstra<sup>748</sup> und verwendet die MQ als inverses Distanzmaß.<sup>749</sup>

### 6.3.3 Auswertungsmöglichkeiten

Die Ergebnisse aller Simulationsläufe werden in CroCoSIM zur Analyse aufbereitet. Dabei steht für jedes generierte Netz eine Übersicht der errechenbaren Eigenschaften zur Verfügung.<sup>750</sup> Sofern eine Batch-Verarbeitung mehrerer Parameter durchgeführt wird, werden die netzbestimmenden Parameter mit den Messgrößen gespeichert.<sup>751</sup>

Mit Bezug auf die Ausführungen zu Verifikation und Validierung von Modell und Instanz wurden zudem grafische Auswertungen zur Sicherstellung der korrekten Evaluation implementiert.<sup>752</sup> Abb. 56 zeigt eine beispielhafte grafische Repräsentation generierter Netze.<sup>753</sup>

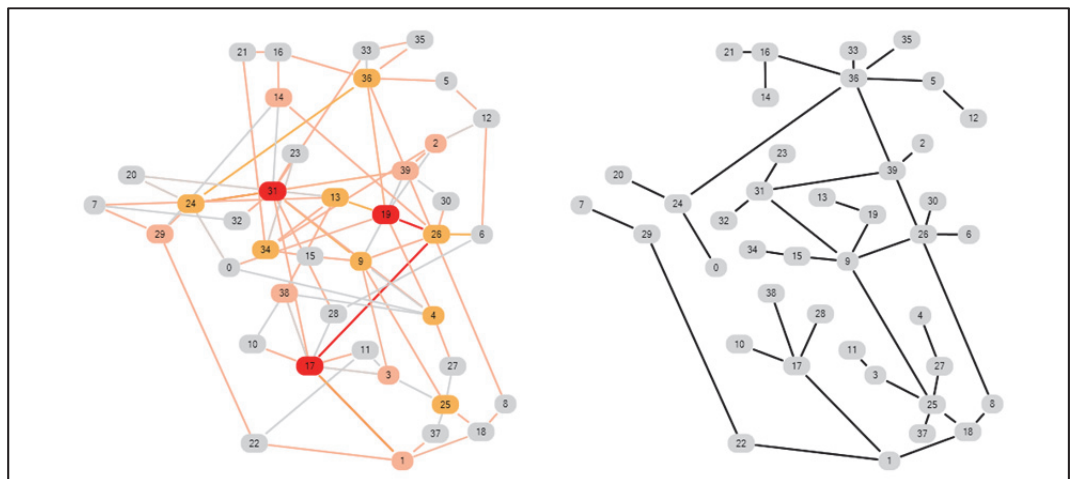


Abb. 56: Grafische Netz- und Routing-Darstellungen in CroCoSIM

(Quelle: Eigene Darstellung)

<sup>748</sup> Für eine Beschreibung des Algorithmus vgl. Gross/Yellen (2006): 179–181. Für eine Beschreibung der Umsetzung vgl. Anhang 1.

<sup>749</sup> Vgl. hierzu auch die Ausführungen zu Metriken und Ähnlichkeitsmaßen in Abschnitt 4.5.2.1.

<sup>750</sup> Vgl. hierzu die Aufzählung der typischen Netzeigenschaften in Abschnitt 5.2.2 sowie die definierten Messgrößen und Robustheitsmaße in Abschnitt 5.3, insbesondere Abschnitt 5.3.4.

<sup>751</sup> Für eine genaue Aufstellung der errechneten Größen und der Repräsentation in der Software vgl. Anhang 1.

<sup>752</sup> BANKS weist darauf hin, dass für Simulationen die grafische (animierte) Darstellung ein sehr gutes Mittel zur Verifikation darstellt (vgl. Banks (1998): 23).

<sup>753</sup> Die dafür verwendete Bibliothek wurde um die Möglichkeit ergänzt, die Darstellungen seed-abhängig zu gestalten, sodass nicht nur die Simulationsläufe, sondern auch die grafischen Abbildungen in CroCoSIM vollständig reproduzierbar sind (vgl. Anhang 1).

Beiden Darstellungen liegt dasselbe skalenfreie Netz mit 40 Teilnehmern zugrunde. Linksseitig wurde statt einer (farb-)neutralen Darstellung eine Einfärbung der Knoten basierend auf ihrem Grad und der Kanten basierend auf der Anzahl von Peer-spezifischen Routings, die diese Kante nutzen, vorgenommen.<sup>754</sup> Rechtsseitig ist das Routing für Peer, 18<sup>6</sup> dargestellt.<sup>755</sup>

Die Darstellung von Knotengraden (vgl. Abb. 57) im Säulendiagramm stellt zudem die Überprüfbarkeit der Einhaltung angenommener Verteilungen sicher, wie sie für skalenfreie Netzwerke (linkes Diagramm,  $n = 200$ ) und Zufallsnetzwerke (mittleres Diagramm,  $n = 200$ ) vorgestellt wurden.<sup>756</sup>

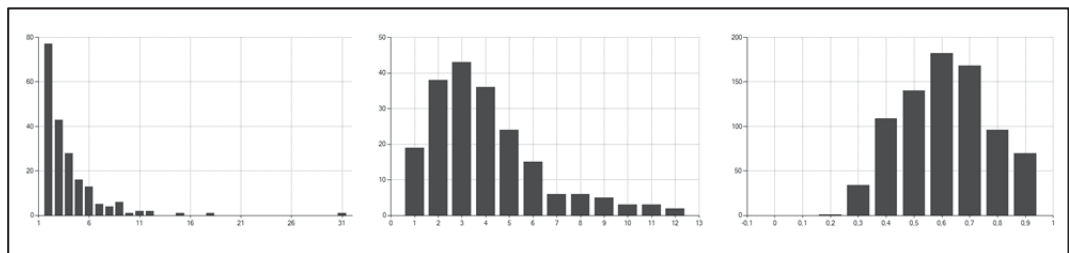


Abb. 57: Diagramm-Darstellungen in CroCoSIM

(Quelle: Eigene Darstellung)

Ergänzend ist im rechten Diagramm ein Histogramm der im Netz tatsächlich ermittelten MQ aufgezeigt, mit dessen Hilfe bestimmt werden kann, welche Parametereinstellungen zu welcher MQ-Verteilung führen, sodass die grafische Auswertung von Extremwert-basierten Einzelsimulationen die Festlegung der Parameter für die Mehrfachsimulation unterstützt.<sup>757</sup>

## 6.4 Durchführung der Simulation und Interpretation der Ergebnisse

Die folgenden Abschnitte stellen die durchgeführten Simulationen und die dafür verwendeten Modelle und deren Parameter vor. Ihre Struktur folgt dabei den Abhängigkeiten zwischen den einzelnen Qualitäten und deren Einflussgrößen.<sup>758</sup> Die

<sup>754</sup> Eine hellgraue Färbung steht für das niedrigste, eine dunkelrote Färbung für das höchste Quartil, hell- und dunkelorange sind entsprechend eingeordnet. Es lässt sich dadurch grafisch sehr leicht die typische Topologie eines skalenfreien Netzwerks erkennen (vgl. Abb. 45).

<sup>755</sup> Die Darstellung einzelner Routings ermöglicht die Verifikation einer eindeutigen (und korrekten) Pfadwahl in der Implementierung.

<sup>756</sup> Vgl. Abschnitt 5.2.2, insbesondere Abb. 46.

<sup>757</sup> Vgl. Abschnitt 6.4.2.

<sup>758</sup> Vgl. Abschnitte 5.2.1 und 5.3.

erzielten Ergebnisse werden zudem im Kontext der unternehmensübergreifenden BI interpretiert. Abschließend erfolgt eine Betrachtung der Zusammenhänge zwischen Mess- und Kontrollgrößen, auf deren Basis geprüft wird, welche Aussagen sich anhand der beobachtbaren Größen über die tatsächlichen Größen treffen lassen. Tab. 16 stellt die durchgeführten Simulationen in einer Übersicht dar. Die detaillierten Parametereinstellungen sind in den jeweiligen Abschnitten aufgeführt, eine abschließende, zusammenfassende Übersicht in Abschnitt 6.4.5.

Abschnitt	Betrachtete Größe(n)	Ziel der Simulation	Verwendete Verfahren
6.4.1	MQ, MQ <sub>s</sub> , MQ <sub>c</sub> , MQ <sub>b</sub>	Ermittlung eines geeigneten Gewichtungsvektors für die Teilqualitäten der MQ	Regressionsanalyse auf Peer-Paarungen; Enumeration von Kombinationen der Teilqualitäten auf den Peer-Paarungen; Rangkorrelation der Ergebnisse
Verwendung des ermittelten Gewichtungsvektors für weitere Analysen			
6.4.2	RQ (NQ), TRQ (TNQ)	Ermittlung einer geeigneten t-Norm zur Errechnung der RQ	Erstellung unterschiedlicher t-Normen; Vergleich in identischen Netzwerken; Rangkorrelation von NQ und TNQ; Bewertung der Güte je t-Norm
Verwendung der ermittelten t-Norm für weitere Analysen			
6.4.3.1	NQ	Ermittlung maßgeblicher Parameter auf die NQ	Enumeration von Kombinationen der möglichen Parameter; Ceteris-paribus-/Ceteris-rectis-Analyse
Einschränkung der Parameter			
6.4.3.2	NQ, TNQ	Überprüfung der Aussagekraft der NQ als Messgröße und Identifikation praxistauglicher Netzwerktypen	Enumeration von Kombinationen der möglichen Parameter; Ceteris-paribus-/Ceteris-rectis-Analyse; Grafische Auswertungen; Hypothesentests
Weitere Einschränkung der Parameter für Detail-Simulationen			
6.4.3.3	NQ, TNQ	Identifikation der einflussreichsten Parameter in den ermittelten Netzwerktypen	Enumeration von Kombinationen der möglichen Parameter; Ceteris-paribus-/Ceteris-rectis-Analyse; Regressionsanalyse unter Berücksichtigung von Interaktionstermen der Parameter
Eingrenzung auf sinnvolle Parameter und Netzwerktypen zur Prognosegütermittlung			
6.4.4	MQ, TMQ, NQ, TNQ (Schätzer)	Ermittlung von Schätzfunktionen für die tatsächlichen Qualitäten	Regressionsanalyse; Mehrstufige Regressionanalyse; Grafische Auswertungen

Tab. 16: Übersicht der durchgeführten Simulationen

(Quelle: Eigene Darstellung)



### 6.4.1 Gestaltung der Mapping-Qualität

Die MQ stellt die verbindungsspezifische Messgröße in einem CroCoBIN dar. Sie wird durch drei Teilgrößen repräsentiert, deren Einflussverhältnis durch einen Gewichtungsvektor  $(\alpha, \beta, \gamma)$  festgelegt wird.<sup>759</sup> Da die Berechnung der MQ notwendige Voraussetzung für die Errechnung der weiteren Messgrößen ist, müssen in einem ersten Schritt geeignete Wertbelegungen für den Gewichtungsvektor bestimmt werden. Zu berücksichtigen ist dabei, dass die MQ die TMQ möglichst gut abbilden sollte, da die MQ in einem CroCoBIN als Messgröße vergleichsweise einfach, die TMQ als Kontrollgröße hingegen nicht oder nur mit erheblichem Aufwand ermittelbar ist.<sup>760</sup> Eine geeignete Wertebelegung für den Gewichtungsvektor liegt daher vor, wenn diese eine der TMQ möglichst proportionale MQ hervorbringt.

Als Basis für die Bestimmung der Wertbelegungen (in Bezug auf die Simulation: Parameterkombinationen) wurden durch vollständige Enumeration aller möglichen Werte für die drei Wahrscheinlichkeiten bei der Erzeugung eines Peerspezifischen Würfels sowie die Verwendung sechs unterschiedlicher Seeds insgesamt etwa 32.000 unterschiedliche Peer-Paarungen erstellt.<sup>761</sup> Um einen ersten Eindruck des Einflusses der einzelnen Parameter auf die Zielgröße zu erhalten, wurde eine multiple Regressionsanalyse mit den erhaltenen Daten durchgeführt. Diese stellt die Abhängigkeit einer Variablen (TMQ) von mehreren gegebenen anderen Variablen ( $MQ_s$ ,  $MQ_c$ ,  $MQ_b$ ) dar und ist im Allgemeinen nützlich, um den funktionalen Zusammenhang der Variablen zu beschreiben, der ohne weitere Anpassung als linear angenommen wird.<sup>762</sup>

Die Anwendung der multiplen Regression führt zu einer Funktionsgeraden der Form:  $TMQ = 1,0639 MQ_s + 0,1013 MQ_c + 0,6578 MQ_b - 1,0192$ . Die Koeffizienten geben dabei jedoch nur an, welche Änderung der TMQ bei einer Änderung einer der unabhängigen Variablen zu erwarten ist. Sie stehen nicht direkt für die Wichtigkeit der einzelnen Variablen und sind darüber hinaus nur vergleichbar,

<sup>759</sup> Vgl. Abschnitt 5.3.1, insbesondere Definition (26).

<sup>760</sup> Vgl. Abschnitt 5.3.1.

<sup>761</sup> Die Enumeration erfolgte dabei in Schritten der Größe 0,1, was für die drei Variablen  $11^3 = 1.331$  Kombinationen erzeugt. Je Seed wurden fünf Peers und vier Verbindungen erstellt, sodass sich eine Gesamtzahl von  $1.331 \times 4 \times 6 = 31.944$  Verbindungen ergibt.

<sup>762</sup> Vgl. Wooldridge (2013): 64–66.



wenn sie in der gleichen Einheit erfasst wurden.<sup>763</sup> Für die vorliegende Problemstellung sind sie daher nicht unmittelbar einsetzbar. Die Koeffizienten können aber, da die Ursprungsvariablen alle per Definition im Wertebereich  $[0; 1]$  liegen,<sup>764</sup> so normiert werden, dass ihre Summe eins ergibt und die MQ als gewichtete Zusammenfassung der einzelnen Teilqualitäten eine linear verschobene und transformierte Einflussgröße für die TMQ abbildet. Es ergeben sich als normierte Werte  $\alpha = 0,5836$ ,  $\beta = 0,0556$  und  $\gamma = 0,3608$ .

Obwohl die Ergebnisse der Regressionsanalyse aus oben genannten Gründen nicht direkt als Gewichtung interpretiert und daher auch nicht in eine solche überführt werden können, lässt sich auf Basis der Analyse annehmen, dass alle Teilqualitäten einen grundsätzlichen Einfluss auf die TMQ aufweisen. Diese Annahme wird durch ein vergleichsweise hohes Bestimmtheitsmaß der Regression sowie durch weitere Relevanztests der Funktion und der einzelnen Koeffizienten gestützt.<sup>765</sup> Eine grafische Auswertung der so gebildeten MQ und der TMQ zeigt allerdings, dass die wichtigste Grundannahme der multiplen Regression – die lineare Abhängigkeit – nicht gegeben ist (vgl. Abb. 58). Die errechneten Gewichtungen können daher nicht ohne ergänzende Überprüfungen für die weiteren Analysen genutzt werden.<sup>766</sup>

---

<sup>763</sup> Vgl. Backhaus (2011): 70.

<sup>764</sup> Vgl. Abschnitt 5.3.1.

<sup>765</sup> Das *Bestimmtheitsmaß*  $R^2 \in [0; 1]$  drückt aus, welcher Anteil der Varianz in den beobachteten Daten durch die gegebenen unabhängigen Variablen erklärt wird (vgl. Backhaus (2011): 72–76). Für die vorliegende Regression ist  $R^2 = 0,8213$ . Die weiteren Tests umfassen einen F-Test zur Bestimmung der Signifikanz (Zutreffen einer Aussage unter Annahme bestimmter Irrtumswahrscheinlichkeiten) der Regression als solcher und t-Tests zur Bestimmung der Signifikanz der einzelnen Koeffizienten. Da die multiple Regression hier nur als Vorstufe für weitere Tests und – aus den im Folgenden erläuterten Gründen – nicht als vollständige Analyse verwendet wird, werden die entsprechenden Begriffe und Vorgehensweisen erst im Zusammenhang tatsächlicher Signifikanzprüfungen von Mess- und Kontrollgrößen diskutiert (vgl. Abschnitt 6.4.3.2; die hier vorgestellten Tests weisen eine Signifikanz mit einer Irrtumswahrscheinlichkeit von  $\alpha = 0,005$ [sic!] aus, wobei  $\alpha$  die statistische Größe (vgl. Wooldridge (2013): 115f.) und nicht den Gewichtungsfaktor der MQ<sub>s</sub> angibt).

<sup>766</sup> Es existieren Methoden zur Abbildung von nicht-linearen Zusammenhängen im Rahmen einer multiplen Regression (vgl. Wooldridge (2013): 66). Diese werden im Zusammenhang der Parameterabhängigkeiten in Netzen und der Prognosefähigkeit der Gütemaße diskutiert (vgl. Abschnitte 6.4.3.3 und 6.4.4), unterstützen jedoch nicht die Suche nach einer geeigneten Parameterkombination des Gewichtungsvektors.

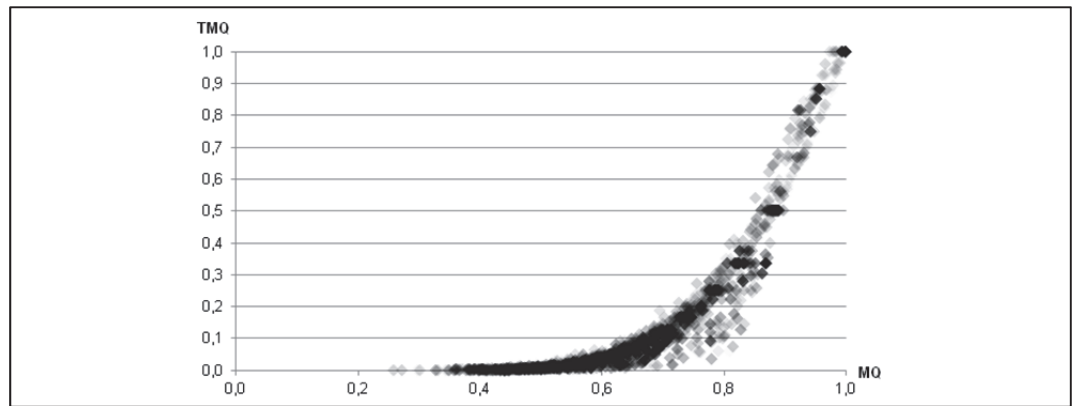


Abb. 58: TMQ-Werte in Abhängigkeit von der MQ mit Gewichtung nach Ergebnissen der multiplen Regression

(Quelle: Eigene Darstellung)

Um die errechneten Werte in ein Verhältnis mit anderen möglichen Gewichtungsvektoren zu setzen, wurden die möglichen Kombinationen der Koeffizienten für  $MQ_s$ ,  $MQ_c$  und  $MQ_b$  verwendet, die sich unter der Bedingung ergeben, dass  $\alpha + \beta + \gamma = 1$  gilt. Die Anzahl der Kombinationen wurde zur Durchführung der Simulation zudem auf 231 beschränkt, indem nur Vielfache von 0,05 als zulässige Werte für die Koeffizienten verwendet wurden.<sup>767</sup> Um die Stärke des Zusammenhangs der so ermittelten MQ-Varianten und der TMQ zu bestimmen, wurde die Korrelation der Werte geprüft. Der für die Bestimmung der Korrelation metrischer Werte eigentlich übliche *Koeffizient von Bravais-Pearson* setzt allerdings – wie die Regressionsanalyse – einen linearen Zusammenhang der Werte voraus.<sup>768</sup> Dieser ist jedoch auch für die Teilqualitäten nicht gegeben (vgl. Abb. 59).

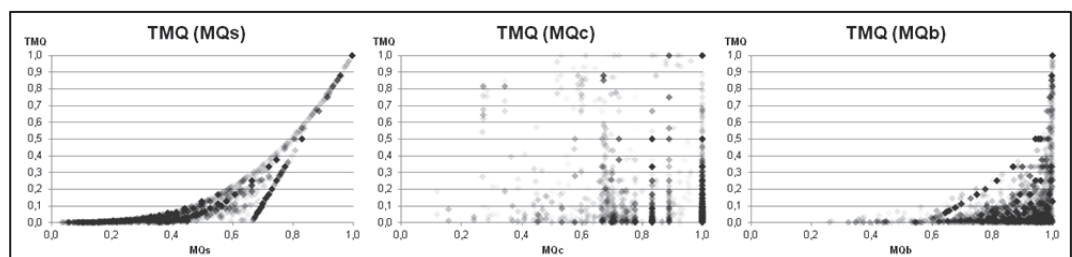


Abb. 59: TMQ-Werte in Abhängigkeit von den Teilqualitäten der MQ

(Quelle: Eigene Darstellung)

<sup>767</sup> Die Reduzierung auf 231 Kombinationen erfolgte vor dem Hintergrund der praktischen Anwendbarkeit in der Simulation und um ein ‚Overfitting‘, d. h. eine Überanpassung des Modells an die Daten, zu vermeiden, das bei der Auswahl zu gezielt auf die Eingabedaten passender Parameter eines Modells bei vielen Analysemethoden auftritt (vgl. Witten/Frank/Hall (2011): 18f., 88).

<sup>768</sup> Vgl. Cleff (2015): 98.

Bei der Betrachtung des visualisierten Zusammenhangs von TMQ und Teilqualitäten zeigt sich zudem, dass die  $MQ_s$  einen ähnlichen Zusammenhang zur TMQ beschreibt wie die oben vorgestellte MQ (vgl. Abb. 58), während die  $MQ_e$  keinen direkt aus dem Diagramm ablesbaren funktionalen Zusammenhang erkennen lässt, was die Annahme unterstützt, dass eine höhere Gewichtung der  $MQ_s$  die Proportionalität zur TMQ begünstigt.<sup>769</sup>

Als rechnerische Alternative kann für nicht-lineare Zusammenhänge der *Korrelationskoeffizient nach Spearman* ( $\rho$ ) (auch: Rangkorrelationskoeffizient) ermittelt werden. Er beschreibt mit  $-1$  einen perfekt negativen, mit  $+1$  einen perfekt positiven und mit  $0$  einen nicht vorhandenen monotonen Zusammenhang der überprüften Wertepaare.<sup>770</sup> Tab. 17 stellt alle Koeffizienten-Kombinationen und die zugehörigen Korrelationskoeffizienten dar, für die  $\rho \geq 0,98$  gilt, d. h. der Zusammenhang nahezu perfekt positiv ist.<sup>771</sup>

<sup>769</sup> Auffällig ist zudem die klar erkennbare ‚Linie‘ im linken Diagramm von den Koordinaten  $[0,67; 0,00]$  zu  $[1,00; 1,00]$ . Die dort erkennbaren Punkte ergeben sich durch die Berechnung der  $MQ_s$  als rein elementbasierter Teilqualität. Werden Würfel miteinander verglichen, die in zwei Dimensionen durch die Belegung der Simulationsparameter zur Element- oder Ebenenauslassung keine Elemente aufweisen, so wird die  $MQ_s$  zwar für drei Dimensionen gebildet, weist jedoch für zwei dieser Dimensionen den Wert 1 auf, da ‚leere‘ Dimensionen vollidentisch sind. Die Simulation bildet damit auch den Praxisfall unvollständiger Datenwürfel von verbundenen Peers ab. Im Sinne der Entwicklung einer robusten MQ (vgl. zu dem Robustheitsbegriff auch die Ausführungen zu Netzwerkeigenschaften in Abschnitt 5.2.2) wurden diese Fälle bewusst mit erzeugt und ausgewertet.

<sup>770</sup> Vgl. Cleff (2015): 102f.

<sup>771</sup> Der minimale Korrelationskoeffizient in der vorliegenden Untersuchung liegt bei  $\rho = 0,0757$ , der entsprechende für den Gewichtungsvektor ist  $(0,00; 1,00; 0,00)$ .

$\alpha$ (MQ <sub>s</sub> )	$\beta$ (MQ <sub>c</sub> )	$\gamma$ (MQ <sub>b</sub> )	$\rho$	
0,60	0,00	0,40	0,9872	
0,65	0,00	0,35	0,9868	
0,60	0,05	0,35	0,9862	
0,55	0,05	0,40	0,9861	
0,55	0,00	0,45	0,9854	
0,70	0,00	0,30	0,9849	
0,65	0,05	0,30	0,9847	
0,55	0,10	0,35	0,9840	
0,50	0,05	0,45	0,9836	
0,60	0,10	0,30	0,9831	
0,50	0,10	0,40	0,9830	
0,75	0,00	0,25	0,9819	
0,70	0,05	0,25	0,9818	
0,50	0,00	0,50	0,9808	
0,65	0,10	0,25	0,9803	
$\varnothing^{772}$	0,60	0,04	0,35	0,9840

Tab. 17: Korrelationskoeffizienten unterschiedlicher MQ-Varianten und TMQ

(Quelle: Eigene Darstellung)

Zu erkennen ist eine starke Ähnlichkeit der Kombinationen, wobei die MQ<sub>s</sub> stets den einflussreichsten Bestandteil darstellt. Der Einfluss der MQ<sub>c</sub> ist durchgängig gering, im Mittel der ‚besten‘ Kombinationen jedoch nicht null.<sup>773</sup> Diese geringe Aussagekraft lässt sich durch die starke Abhängigkeit der Konsistenzbedingung von dem Vorhandensein von Hierarchiewechseln erklären, wie sie im vorliegenden Modell unter anderem zur Abbildung von Elternelementwechselln eingesetzt werden.<sup>774</sup> Sofern eine alternative Hierarchie verwendet wird, wechselt das Elternelement ggf. für eine sehr große Anzahl an Elementen, sodass selbst bei starker Aufrechterhaltung der Soundness einer Überführung die Konsistenzbedingung für die betrachtete Dimension nur schwach erfüllt wird.<sup>775</sup> Die ermittelten Koeffizien-

<sup>772</sup> Die Abweichung der Summe der Durchschnittswerte der angegebenen 15 Einzelkombinationen von 1,0 ergibt sich durch Rundungsdifferenzen. Die exakten Werte sind 0,60 $\bar{3}$ ; 0,04 $\bar{3}$ ; 0,35 $\bar{3}$ .

<sup>773</sup> Die vorgestellten Analysen basieren auf der Betrachtung und Verwendung von Durchschnittswerten. Dies ist grundsätzlich angemessen und empfehlenswert, um die beschriebene Überanpassung an Modelldaten zu vermeiden. Ein analoges Vorgehen wird auch im Bereich anderer Datenanalysen und Klassifikationsermittlungen eingesetzt (vgl. Freund/Mansour/Schapire (2001): 1f. Die Autoren weisen nach, dass die Betrachtung von Durchschnittswerten in Klassifikationsalgorithmen vergleichbar gut zur besten Einzelhypothese ist, dabei aber deutlich stabiler.).

<sup>774</sup> Vgl. Abschnitt 6.3.2.1.

<sup>775</sup> Es liegt allerdings keine Abhängigkeit der beiden Teilqualitäten vor ( $\rho_{MQ_s, MQ_c} = 0,1233$ ), sodass die erzielten Ergebnisse für alle Teilqualitäten aussagekräftig sind.

ten sind daher immer vor dem Hintergrund der Modellspezifika zu interpretieren.<sup>776</sup>

Die den Durchschnittswerten nächste Parameterkombination mit Vielfachen von 0,05 ist (0,60; 0,05; 0,35). Sie entspricht damit auch der Parameterkombination, die durch die multiple Regression bestimmt wurde.<sup>777</sup> Zur Bestätigung der Ergebnisse wurde eine weitere Simulation durchgeführt, in der die Anzahl der erzeugten Verbindungen deutlich auf 1,02 Millionen erhöht wurde. Die Ergebnisse einer auf diesen Verbindungsdaten durchgeführten multiplen Regression – zur Überprüfung der Gleichartigkeit beider betrachteter Datenreihen – entsprechen derselben Parameterkombination. Die Güte der entsprechenden Rangkorrelation liegt bei  $\rho = 0,9828$ . Im Rahmen des vorliegenden Modells wird daher für alle nachfolgenden Simulationen die MQ durch einen Gewichtungsvektor der Teilqualitäten errechnet als  $MQ = 0,60 MQ_s + 0,05 MQ_c + 0,35 MQ_b$ .

#### 6.4.2 Auswahl geeigneter Kombinationsverfahren zur Bestimmung der Routing-Qualität

Die Ermittlung geeigneter Werte für die Berechnung der MQ ermöglicht in einem nächsten Schritt die Prüfung von Verbindungskombinationen, d. h. Pfaden sowie deren Kombinationen zu Routings. Das Ziel ist dabei die Auswahl einer geeigneten t-Norm zur weiteren Untersuchung ganzer Netzwerke. Diese soll zum einen eine möglichst präzise Übereinstimmung, bzw. hohe Proportionalität, von Messgrößen (PQ, RQ, NQ) und Kontrollgrößen (TPQ, TRQ, TNQ) ermöglichen, zum anderen eine hohe tatsächliche Qualität der Netzwerke begünstigen. Sie muss also zu einer Auswahl möglichst geeigneter Pfade und Routings führen. Die beiden Ziele sind dabei nicht zwangsläufig vollständig kongruent.<sup>778</sup>

<sup>776</sup> Eine Betrachtung alternativer Modellspezifikationen oder Realweltphänomene (im vorliegenden Fall z. B. der eingesetzten Matching-Verfahren) und deren möglicher Auswirkungen auf die hier vorgestellten Betrachtungen und Ergebnisse wird in Abschnitt 7.2 vorgenommen.

<sup>777</sup> Berücksichtigt werden dabei ausgehend von den exakten Werten der multiplen Regression wiederum die Vielfachen von 0,05.

<sup>778</sup> Denkbar ist, dass eine t-Norm z. B. immer ‚schlechte‘ Routings auswählt, diese aber präzise bewerten kann, sodass sie eine sehr hohe Korrelationen von Mess- und Kontrollgröße aufweist. Tatsächlich trifft dies – in Vorwegnahme der weiteren Ergebnisse – insbesondere auf die Minimum-Norm zu.

Um eine hohe Anzahl an Routings zu prüfen, wurden Netzwerke unterschiedlicher Größenklassen<sup>779</sup> mit variierenden Netzparametern erzeugt, sodass exakt 15,825 Mio. unterschiedliche Pfade zur Untersuchung verwendet werden konnten.<sup>780</sup> Für jeden Pfad – und damit auch für alle Routings und Netzwerke – wurden sowohl die Messgröße als auch die Kontrollgröße unter Verwendung 35 unterschiedlicher t-Normen ermittelt.<sup>781</sup> Zur Auswertung wurde auf die NQ und TNQ zurückgegriffen, die eine aggregierte Darstellung der Ergebnisse ermöglichen. Zudem spiegeln diese Qualitäten den Effekt auf das ganze Netzwerk wider, was die Aussagekraft der Untersuchung erhöht, da eine isolierte Bestimmung der Pfad- und Routingqualitäten ignorieren würde, dass in einem Netzwerk eine einheitliche t-Norm verwendet werden muss und die ermittelten Routings alle dieselben Verbindungen als Rechenbasis verwenden. Abb. 60 stellt beispielhaft eine Übersicht aller ermittelten NQ/TNQ-Kombinationen für  $T_1^H$  (das algebraische Produkt) dar (oben) und spaltet diese auch nach den unterschiedlichen Größenklassen der Netzwerke auf (unten). Anhand der grafischen Darstellung ist dabei zunächst kein eindeutiger funktionaler Zusammenhang der Variablen zu erkennen, was möglicherweise auch durch den hier nicht transparenten Einfluss weiterer Parameter zu erklären ist. Die Unterscheidung nach Größenklassen scheint jedoch auf Grund der deutlich erkennbaren Konzentrationen in unterschiedlichen Achsenabschnitten gerechtfertigt.<sup>782</sup>

<sup>779</sup> Es wurden vier Größen verwendet, mit  $n \in \{25; 50; 100; 200\}$ . Diese vier Klassen bilden typische Größen von CroCoBIN-Netzwerken ab (vgl. Abschnitt 6.2) und ermöglichen (im Gegensatz zu randomisierten Netzwerkgrößen) die Analyse von Netzwerkeigenschaften in Abhängigkeit von der Teilnehmerzahl. Da die Bestimmung der Kombinationsfunktion der Netzwerkanalyse vorgeht, lassen sich hier noch keine weiteren Aussagen über die unterschiedlichen Einflüsse der weiteren Parameter treffen. Sie werden ausführlich in Abschnitt 6.4.3.1 diskutiert.

<sup>780</sup> Die angegebene Zahl beschreibt nur die tatsächlich in den ermittelten Routings verwendeten Pfade. Die Anzahl potenzieller und damit auch bei der Routerstellung geprüfter Pfade ist erheblich größer.

<sup>781</sup> Betrachtet wurden die in Abschnitt 5.3.2 vorgestellten Hamacher-, Frank- und Yager-Normen  $T_\lambda^H$ ,  $T_\lambda^F$ ,  $T_\lambda^Y$  mit  $\lambda \in \{0,00; 0,25; 0,50; 0,75; 1,00; 1,25; 1,50; 1,75; 2,00; 3,00; 5,00; 10,00\}$ , wobei  $T_1^F = T_1^H$  nur einmal ermittelt wurde und  $T_0^F$  die Minimum-Norm abbildet.

<sup>782</sup> Eine ausführliche Analyse der weiteren Parameter ist hier aus oben genannten Gründen nicht möglich. Vorweggenommen in Bezug auf die Einflussgrößenbestimmung unterschiedlicher Parameter (vgl. Abschnitt 6.4.3.1) sei aber, dass der dort identifizierte maßgebliche Einfluss des durchschnittlichen Knotengrades hier *nicht* zum Tragen kommt. Die sichtbare ‚Spreizung‘ der Datenreihen ist vielmehr auf den ‚selfish‘-Parameter zurückzuführen. Sie wird zunächst nicht weiter betrachtet, da die zu identifizierende Norm alle Netzarten und -größen möglichst zutreffend abbilden soll, um eine möglichst ausgewogene Analyse vollständiger Netzwerke zu ermöglichen.

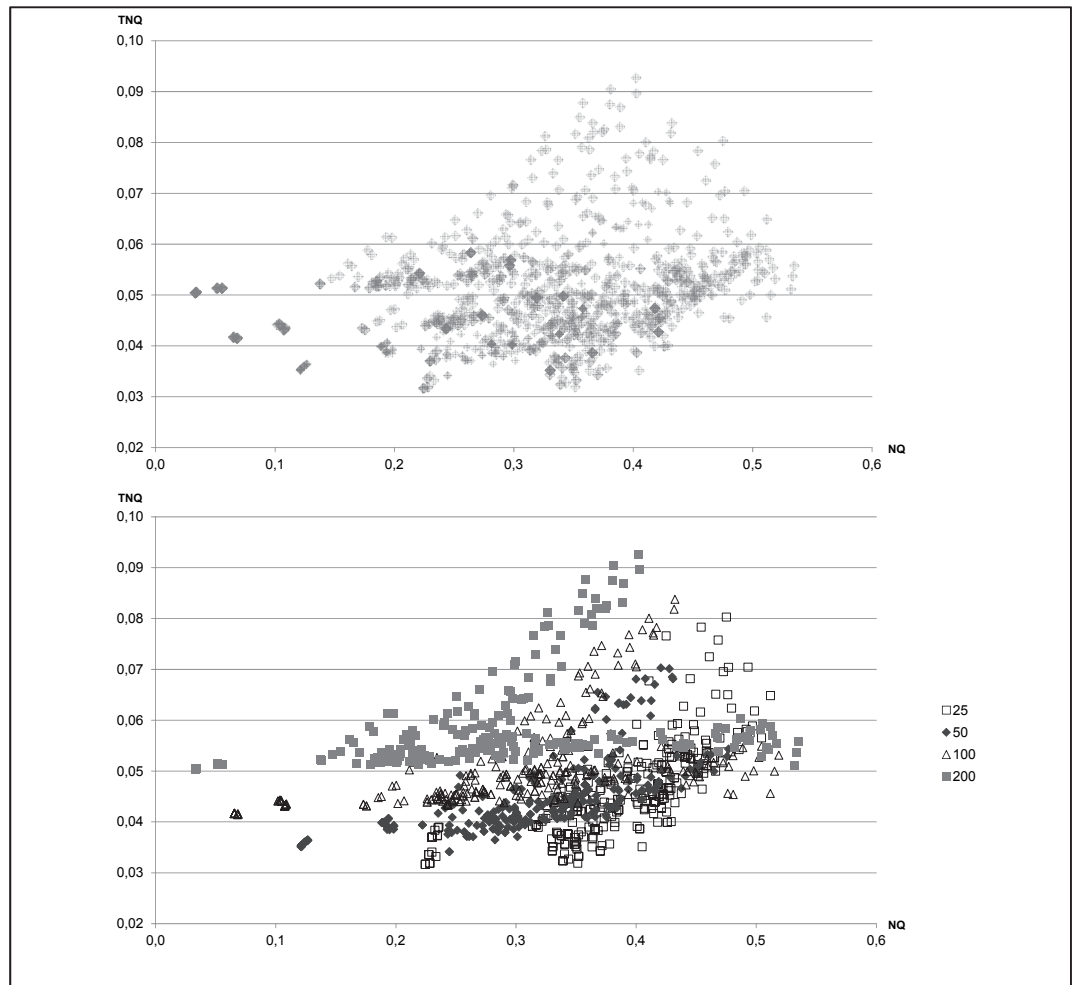


Abb. 60:  $TNQ$  in Abhängigkeit von der  $NQ$  sowie  $NQ$  und Netzwerkgröße unter Verwendung von  $T^H_1$

(Quelle: Eigene Darstellung)

Ein linearer Zusammenhang ist – anders als bei der Betrachtung von  $MQ$  und  $TMQ$  –<sup>783</sup> nicht eindeutig auszuschließen, sodass die Ermittlung eines Korrelationskoeffizienten nach Bravais-Pearson für alle  $t$ -Normen grundsätzlich möglich ist. Allerdings weisen die Werte der  $NQ$  bereits für die vorgestellte  $T^H_1$ -Norm keine Normalverteilung auf, was die Aussagekraft eines möglichen linearen Zusammenhangs auf den Stichprobenumfang einschränkt und keine Rückschlüsse auf weitere Netzwerke zulässt.<sup>784</sup>

<sup>783</sup> Vgl. Abb. 58.

<sup>784</sup> Vgl. zu den Eigenschaften und Voraussetzungen der Korrelationskoeffizienten Bortz/Schuster (2010): 161f. Die Überprüfung der Normalverteilung wurde mit einem Shapiro-Wilk-Test durchgeführt, der als Standardtest für diese Überprüfungen gelten kann. Im Vergleich zu alternativen Testmethoden weist er empirisch nachweisbar eine von der Art der tatsächlichen vorliegenden Verteilung unabhängige hohe Güte auf (vgl. Yazici/Yolacan (2007): 182f.).

Für alle ermittelten Kombinationen wurde daher der Rangkorrelationskoeffizient bestimmt. Des Weiteren wurde für alle t-Normen die durchschnittliche TNQ errechnet, um alle t-Normen hinsichtlich der oben vorgestellten Ziele bewerten zu können. Die Ergebnisse wurden dabei für alle Netzgrößen errechnet und im Anschluss standardisiert (z-transformiert<sup>785</sup>), um eine Vergleichbarkeit der Werte sicherzustellen. Die standardisierten Werte wurden im Anschluss ungewichtet gemittelt und die geprüften t-Normen entsprechend des erreichten Wertes sortiert.<sup>786</sup> Tab. 18 stellt die Reihenfolge aller t-Normen in Bezug auf die Korrelationsgüte von NQ und TNQ, die durchschnittlich erzielte TNQ und den Mittelwert der beiden Größen dar. Zusätzlich sind der standardisierte Wert der Gesamtbeurteilung sowie die erzielten Ränge in Bezug auf den Wert je Netzwerkgröße angegeben. Hervorgehoben sind die ‚beste‘ Norm in Bezug auf den Gesamtwert  $T_{1,25}^Y$ , die Minimum-Norm  $T_0^F$  und das algebraische Produkt  $T_1^H$ .

---

<sup>785</sup> Eine *z-Transformation* entsteht durch  $z = \frac{x - \bar{x}}{\sigma}$  und ermöglicht den Vergleich von Werten unterschiedlicher Bereiche (vgl. Bortz/Schuster (2010): 35f.). Eine solche Transformation ist notwendig, da der Korrelationskoeffizient  $\rho$  (bei positivem Zusammenhang) zwar so wie das arithmetische Mittel der TNQ aufgrund deren definiertem Wertebereich im Intervall  $[0; 1]$  liegt, die Einheit und Aussage der Werte jedoch differieren.

<sup>786</sup> Eine – zur Plausibilisierung durchgeführte – Gewichtung der Werte in beide Richtungen bis zu einem Verhältnis von 2:1, bzw. 1:2 verändert die Reihenfolge nur minimal, die im Folgenden vorgestellten Schlussfolgerungen überhaupt nicht.



Rang	Reihenfolge Korrelation	Reihenfolge Mittelwert TNQ	Reihenfolge Gesamt	z-Wert Gesamt	Rang $n = 25$	Rang $n = 50$	Rang $n = 100$	Rang $n = 200$
1	Yag(5.000)	Ham(5.000)	Yag(1.25)	0,3904	6	3	2	1
2	Yag(10.000)	Ham(10.000)	Yag(1.000)	0,3863	3	2	26	24
3	Fra(0.000)	Ham(3.000)	Ham(5.000)	0,3219	4	4	27	25
4	Yag(3.000)	Fra(10.000)	Ham(10.000)	0,3088	1	1	28	27
5	Yag(2.00)	Yag(1.00)	Fra(10.00)	0,3043	7	6	22	15
6	Ham(0.00)	Ham(2.00)	Ham(3.00)	0,2975	5	5	25	19
7	Yag(1.75)	Fra(5.00)	Ham(2.00)	0,2924	8	7	23	21
8	Yag(1.25)	Ham(1.75)	Fra(1.50)	0,2914	15	15	15	3
9	Ham(0.25)	Fra(3.00)	Fra(3.00)	0,2899	11	9	19	11
10	Yag(1.50)	Ham(1.50)	Fra(2.00)	0,2889	12	12	17	18
11	Yag(1.00)	Fra(2.00)	Fra(1.75)	0,2877	14	13	14	14
12	Fra(0.25)	Yag(1.25)	Fra(5.00)	0,2876	9	8	20	17
13	Ham(0.50)	Fra(1.75)	Ham(1.50)	0,2870	13	10	18	16
14	Fra(0.50)	Ham(1.25)	Yag(1.50)	0,2867	26	16	1	2
15	Ham(0.75)	Fra(1.50)	Fra(1.25)	0,2864	17	17	13	4
16	Fra(0.75)	Fra(1.25)	Ham(1.75)	0,2863	10	11	21	13
17	Ham(1.00)	Ham(1.00)	Ham(1.25)	0,2838	16	14	16	12
18	Fra(1.25)	Fra(0.75)	Fra(0.75)	0,2801	19	19	11	6
19	Fra(1.50)	Ham(0.75)	Ham(1.00)	0,2796	20	18	12	5
20	Fra(1.75)	Fra(0.50)	Ham(0.75)	0,2737	21	20	10	8
21	Ham(1.25)	Yag(0.75)	Fra(0.50)	0,2711	22	22	9	7
22	Fra(2.00)	Ham(0.50)	Ham(0.50)	0,2622	24	23	8	9
23	Fra(3.00)	Yag(1.50)	Ham(0.00)	0,2554	18	26	3	23
24	Ham(1.50)	Fra(0.25)	Fra(0.25)	0,2534	25	24	7	10
25	Fra(10.00)	Ham(0.25)	Ham(0.25)	0,2529	23	25	6	22
26	Ham(5.00)	Ham(0.00)	Yag(1.75)	0,1996	27	27	5	20
27	Ham(2.00)	Yag(1.75)	Yag(0.75)	0,1606	2	21	30	29
28	Ham(1.75)	Yag(0.50)	Yag(2.00)	0,1124	29	28	4	26
29	Fra(5.00)	Yag(2.00)	Yag(3.00)	-0,1466	30	29	24	28
30	Ham(3.00)	Yag(0.25)	Yag(5.00)	-0,4801	32	31	29	30
31	Ham(10.00)	Yag(0.00)	Yag(0.50)	-0,8481	31	30	33	32
32	Yag(0.75)	Yag(3.00)	Yag(10.00)	-0,9365	33	32	31	31
33	Yag(0.50)	Yag(5.00)	Fra(0.00)	-1,2322	34	33	32	33
34	Yag(0.25)	Yag(10.00)	Yag(0.25)	-2,0619	28	34	34	34
35	Yag(0.00)	Fra(0.00)	Yag(0.00)	-2,0729	35	35	35	35

Tab. 18: Auswertung unterschiedlicher  $t$ -Normen

(Quelle: Eigene Darstellung)

Die Auswertung zeigt, dass  $T_{1,25}^Y$  mit geringem Abstand zu  $T_{1,00}^Y$ , aber bereits deutlichem Abstand zu der drittbesten Norm  $T_{5,00}^H$ , in Bezug auf die z-Werte die Auswahl der  $t$ -Normen dominiert. Sie ist zudem unter den hochbewerteten Normen die einzige, die für alle Netzgrößen gute Resultate erzielt und auch bei größe-

ren Netzen vorteilhaft einzusetzen ist. Deutlich zu erkennen ist die eingeschränkte Verwendbarkeit anderer Normen, die nur in bestimmten Größenklassen gute Resultate erzielen, insbesondere  $T_{10,00}^H$ , die für kleine Netze sehr gut geeignet ist, bei Netzgrößen von  $n \geq 100$  jedoch von mehr als zwei Drittel der anderen Normen übertroffen wird. Eine umgekehrte Aussage trifft auf  $T_{1,50}^Y$  zu, was bei nur geringer  $\lambda$ -Abweichung zu  $T_{1,25}^Y$  bestätigt, wie entscheidend die korrekte Parametrisierung der Normen ist.<sup>787</sup> Die Ausgewogenheit von  $T_{1,25}^Y$  zeigt sich weiterhin daran, dass sie bei Gewichtung der beiden Zielgrößen den ersten Rang einnimmt, jedoch in keiner der beiden Größen bei Einzelbetrachtung zu den ersten fünf Normen gehört. Die Problematik der Einzelwertungen wird dabei besonders deutlich bei der Minimum-Norm  $T_{0,00}^F = T_M$  und den ‚extrem‘ parametrisierten  $T_{5,00}^Y$  und  $T_{10,00}^Y$ . Diese bilden zwar eine sehr starke Korrelation von NQ und TNQ ab, erzielen diese aber nur durch eine korrekte Bewertung und Auswahl schlechter Pfade im Netz, die sie auf Basis ihrer spezifischen Errechnung bevorzugen.<sup>788</sup> Klar erkennbar ist auch die nur durchschnittliche Qualität des algebraischen Produkts als t-Norm, wie es MANDREOLI ET AL. in semantischen P2P-Netzen einsetzen.<sup>789</sup> Allen ‚guten‘ t-Normen ist die Eigenschaft gleich, dass sie, wenngleich sie nicht vollständig konvex sind, sehr niedrige Werte aufweisen, solange die Summe der Eingangswerte ungefähr kleiner eins ist ( $T(x, y) \ll x$  und  $T(x, y) \ll y$ , wenn  $x + y \lesssim 1$ ).

<sup>787</sup> Vgl. hierzu auch die Hinweise zu veränderten Rangreihenfolgen der Normen gleicher Familien in Abschnitt 5.3.2.

<sup>788</sup> Vgl. für eine detaillierte Beschreibung der einzelnen Normen Abschnitt 5.3.2, insbesondere Abb. 50.

<sup>789</sup> Vgl. Mandreoli et al. (2006): 21.

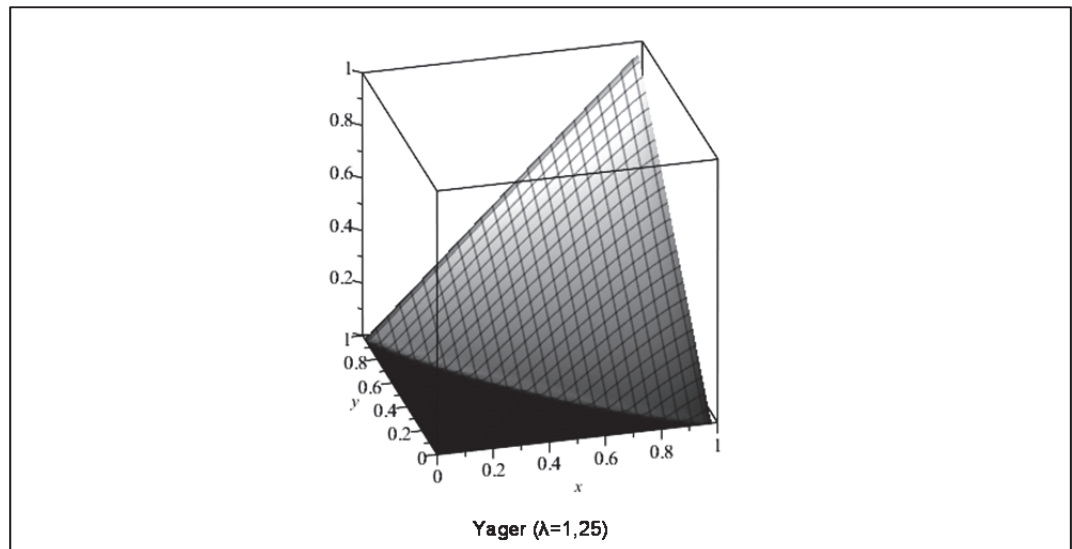


Abb. 61: 3D-Plot der  $t$ -Norm  $T_{1,25}^Y$

(Quelle: Eigene Darstellung)

Abb. 61 zeigt die 3-dimensionale Ansicht der führenden Norm  $T_{1,25}^Y$ . Deutlich erkennbar ist der als schwarze Fläche dargestellte Bereich der Kombination geringer Eingangswerte, für die  $T_{1,25}^Y(x, y) = 0$  gilt. Dies entspricht bei der Verwendung zur Pfadbildung einer starken ‚Bestrafung‘ von schlechten Verbindungen, wie sie z. B. bei einer Minimum-Norm nicht oder nur sehr eingeschränkt stattfindet. Insbesondere in längeren Pfaden können sehr gute Verbindungen damit schlechte Teilstücke nicht mehr ‚kompensieren‘. Dieses Verhalten ist auf Basis der bisherigen Betrachtung als wünschenswert zu beurteilen.<sup>790</sup> Es erklärt zudem das zunehmend schlechte Abschneiden einzelner Normen bei der Vergrößerung der Netze. In größeren Netzen werden üblicherweise längere Pfade notwendig,<sup>791</sup> sodass schlechte Teilstücke durch Normen, die zwar niedrige Werte, aber nicht Nullwerte für niedrige Eingangswerte aufweisen – wie es z. B. beim algebraischen Produkt  $T_{1,00}^H$  der Fall ist, nur unzureichend ‚bestraft‘ werden und keine dauerhaft gute Pfadselektion stattfindet. Auf Basis der durchgeführten Untersuchungen wird für die weiteren Simulationen daher einheitlich  $T_{1,25}^Y$  verwendet.

<sup>790</sup> Vgl. Abschnitte 5.3.2 und 5.3.3.

<sup>791</sup> Vgl. Abschnitt 6.4.3.

### 6.4.3 Betrachtung unterschiedlicher Netzstrukturen, Rahmenbedingungen und Strategien zur Nachbarschaftswahl

Die Festlegung des Gewichtungsvektors der MQ und die Auswahl einer t-Norm sind notwendige Voraussetzungen zur einheitlichen und vergleichbaren Simulation von Netzen, da sie die Errechnung der NQ maßgeblich beeinflussen. Auf Basis der erlangten Erkenntnisse wurden nachfolgend Tests mit stark unterschiedlichen Parameterkombinationen zur Netzgenerierung durchgeführt, um den Einfluss einzelner oder kombinierter Parameter festzustellen, diese in Bezug zu der TNQ zu setzen und im Anschluss empfehlenswerte Parameterkombinationen für CroCoBIN-basierte Netzwerke herauszustellen.

#### 6.4.3.1 Übergreifende Parameteranalyse

Die nachfolgenden Untersuchungen basieren auf 85.008 Netzen, deren Parameterkombinationen für die vorgestellten Größenklassen einen möglichst großen Wertebereich abdecken. Tab. 19 stellt die verwendeten Parameterbereiche der Simulation dar.

Parameter	Wertebereich
Basic algorithm	Random; Scale-free
Net size	25;50;100;200
Target degree	[4;10] @ 1 <sup>792</sup>
m	[2;5] @ 1
Core net size	1; 5
Seed	[1;3] @ 1
PMQG	[0,0;1,0] @ 0,1
Target degree spread	[0,0;1,0] @ 0,1
Sorted node list	True; False
$\mu$	[0,00; 5,00] @ 0,25
Selfish approach	True; False
MQ ( $\alpha$ , $\beta$ , $\gamma$ ) <sup>793</sup>	(0,60; 0,05; 0,35)

Tab. 19: Parameterkonfigurationen der ersten Netzwerk-Simulationsreihe  
(Quelle: Eigene Darstellung)

<sup>792</sup> Das @-Zeichen gibt die Schrittweite im Intervall an und wird hier zur Vermeidung einer umständlichen, parameterspezifischen Schreibweise (wie z. B. 1|deg, 1|mu, ...) vereinfachend eingesetzt.

<sup>793</sup> Die Erzeugung der Datenwürfel in den Peers wurde mit den Werten (0,25; 0,60; 0,50) für die vorgestellten Wahrscheinlichkeiten durchgeführt, was eine PMQG-abhängige, (annähernd) normalverteilte MQ-Verteilung in allen fertigberechneten Netzen verursacht, sodass unterschiedliche Pfadqualitäten für die Ermittlung der Routings sichergestellt sind.

Für jede mögliche Kombination wurde die NQ des entsprechenden Netzes errechnet. Sie bildet die für die Netzteilnehmer ermittelbare Messgröße ab und steht daher zunächst im Fokus bei der Identifizierung einflussreicher Parameter(größen).<sup>794</sup> Für diese werden ceteris-paribus-Analysen<sup>795</sup> durchgeführt. Abb. 62 zeigt für alle nicht-algorithmusspezifischen<sup>796</sup> Parameter die niedrigste und höchste durchschnittliche NQ bei Auswahl der entsprechenden Parameterbelegung.<sup>797</sup>

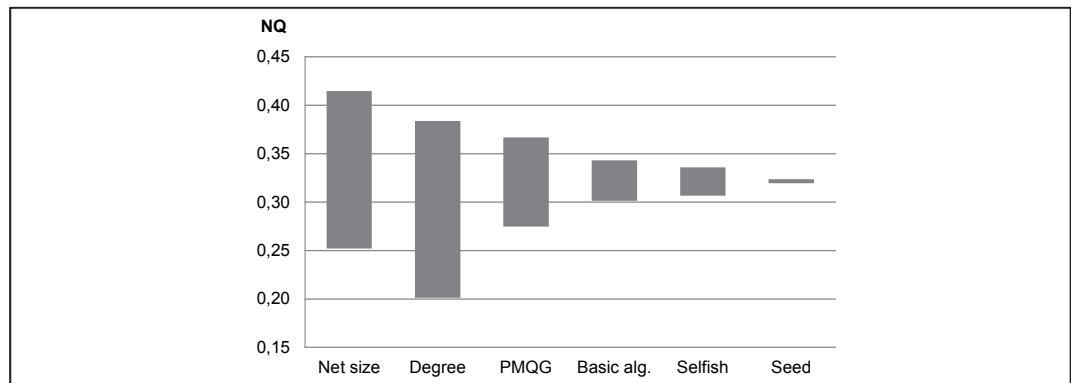


Abb. 62: Wertebereiche der durchschnittlichen NQ in Abhängigkeit von einzelnen Parametern

(Quelle: Eigene Darstellung)

Deutlich erkennbar sind der hohe Einfluss von Netzwerkgröße und Gradanzahl auf die NQ. Zudem weist die Gradanzahl einen höheren Ausschlag in den niedrigen Bereich auf als die anderen Werte, die positiv und negativ gleichmäßig um den durchschnittlichen NQ-Wert über alle Netze von 0,3213 verteilt sind. Dies erklärt sich durch die geringen Werte, die bei sehr niedriger Gradanzahl erzielt

<sup>794</sup> Der Vergleich zur TNQ wird im nachfolgenden Abschnitt vorgenommen. Basierend auf den Erkenntnissen zur Ermittlung geeigneter t-Normen ist eine netzgrößenspezifische, starke Korrelation von NQ zu TNQ anzunehmen (im Mittel:  $\rho = 0,7852$ ), sodass die Verwendung der Messgröße bei der initialen Beobachtung im Vergleich zur Verwendung der Kontrollgröße keinen relevanten Nachteil vermuten lässt.

<sup>795</sup> Eine *ceteris-paribus-Analyse* untersucht den Einfluss einer bestimmten Variablen auf die beobachtete Zielgröße unter der Annahme, dass alle anderen beeinflussenden Variablen gleich bleiben (vgl. Schurz (2014): 1801f.) Die durchgeführten Simulationen liefern entsprechende Ergebnisse. Es werden hier allerdings zur Abbildung größerer Datenmengen nicht die Einzeldatensätze mit Veränderung nur eines Parameters, sondern die Menge aller Datensätze unter Betrachtung des Durchschnitts der übrigen Parameter betrachtet, sodass sich ggf. nur Aussagen ableiten lassen, die ausschließlich in einem bestimmten Spektrum Gültigkeit besitzen. SCHURZ bezeichnet dies abgrenzend als *ceteris-rectis-Analyse* (vgl. Schurz (2014): 1803).

<sup>796</sup> Der Parameter ‚Degree‘ bildet für Zufallsnetzwerke den ‚Target Degree‘, für skalenfreie Netzwerke den Parameter ‚m‘ ab. Vgl. für den Zusammenhang dieser Tab. 14.

<sup>797</sup> Es wurde dafür eine vollständige Iteration aller Parameter durchgeführt und der Wert des Parameters verwendet, der zur niedrigsten bzw. höchsten NQ führt. Dieser entspricht nicht notwendigerweise dem niedrigsten/höchsten Wert im Wertebereich des Parameters.

werden, welche insbesondere in großen Netzen zu sehr langen Pfaden und damit i. d. R. zu schlechten Pfad-Qualitäten führt. Für die ceteris-paribus-Analyse ist die Gradanzahl dennoch ein entscheidender Parameter. Die Auswertung erfolgt dabei – wie bereits in Abschnitt 6.4.2 – nach den vorgestellten Größenklassen. Diese weisen einen erheblichen Einfluss auf die NQ auf und sind – anders als die Gradanzahl und die kleinschrittig überprüften Parameter – aufgrund ihrer großen Wertunterschiede und Interdependenzen nicht fehlerfrei als ceteris-paribus-Analyse zu interpretieren.<sup>798</sup> Zudem hängt die Dichte des Netzwerks, die einen erheblichen Einfluss auf die Robustheit und auf vermutete NQ-Änderungen hat,<sup>799</sup> sowohl von Größe als auch Gradanzahl des Netzwerks ab. Ein (unabhängiger) Effekt kann daher bei Durchschnittsbildung erheblich unterschiedlicher Netzgrößen nicht zweifelsfrei betrachtet werden. Des Weiteren erschwert die Gesamtbeachtung aller Netzwerkgrößen auch Rückschlüsse auf die TNQ, da – wie in Abschnitt 6.4.2 gezeigt – die Rangkorrelation innerhalb einer Netzgröße sehr hoch ist, jedoch bei gleichzeitiger Betrachtung aller Datenpunkte erheblich sinkt ( $\rho = 0,2174$ ).<sup>800</sup> Die unterschiedliche NQ-Verteilung zeigt sich auch direkt bei der Aufschlüsselung der Netzwerke nach dem NQ-basierten Anteil an der Grundgesamtheit der entsprechenden Größenklasse (vgl. Abb. 63).

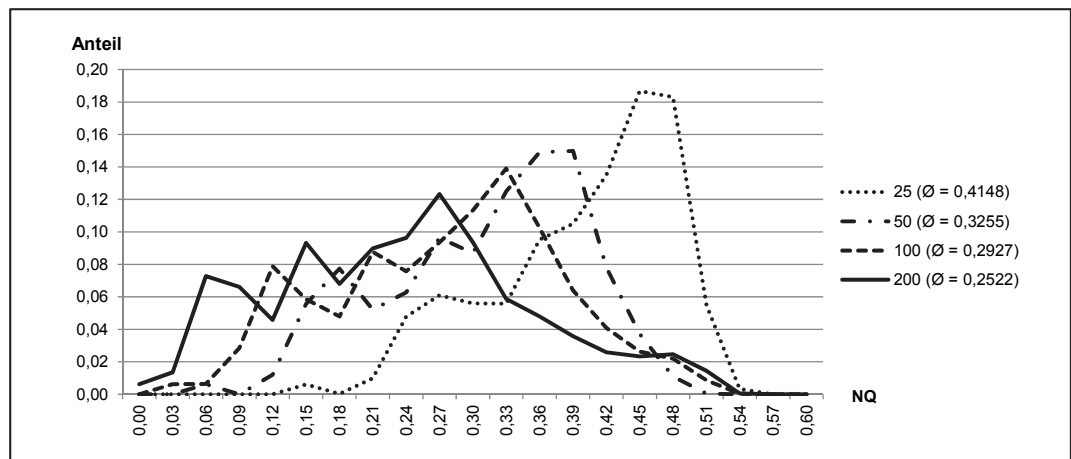


Abb. 63: Anteil Netzwerke mit gegebener NQ nach Größenklassen

(Quelle: Eigene Darstellung)

<sup>798</sup> Die Parameterbelegungen der Netzwerkgröße stellen jeweils eine Verdopplung des vorherigen Wertes dar und können daher nicht als ‚kleine Änderung‘ bei ansonsten gleicher Netzwerkkonfiguration betrachtet werden.

<sup>799</sup> Vgl. Abschnitte 5.3.4 und 5.4.

<sup>800</sup> Die verwendete t-Norm  $T_{1,25}^Y$  weist unter den ersten sieben Normen aber auch für diese Berechnung die höchste Korrelation auf und wird erst von  $T_{1,50}^F$  ( $\rho = 0,2345$ ) übertroffen.

Kleinere Netzwerke führen zu einer durchschnittlich höheren NQ (vgl. die Legende in Abb. 63) und konzentrieren sich auch stärker in den höheren NQ-Bereichen. Es ist damit keine direkte Vergleichbarkeit der Größenklassen gegeben. Aus anwendungsorientierter Sicht lässt sich zudem ergänzen, dass es für tatsächlich implementierte CroCoBIN-Netzwerke aufgrund des üblicherweise in gewissem Rahmen vorgegebenen Teilnehmerkreises einfacher sein dürfte, die Anzahl der Verbindungen je Teilnehmer als die Anzahl der Teilnehmer im Netz maßgeblich zu verändern.

Für den Vergleich der Netzwerke unter veränderter Gradanzahl wurden die Zufallsnetzwerke mit *Target degree*  $\in [4; 6; 8; 10]$  und die skalenfreien Netzwerke mit  $m \in [2; 3; 4; 5]$  herangezogen, die zu (nahezu) identischen Netzdichten führen.<sup>801</sup> Abb. 64 zeigt die entsprechende Auswertung nach Netztypen (Basisalgorithmen) getrennt.

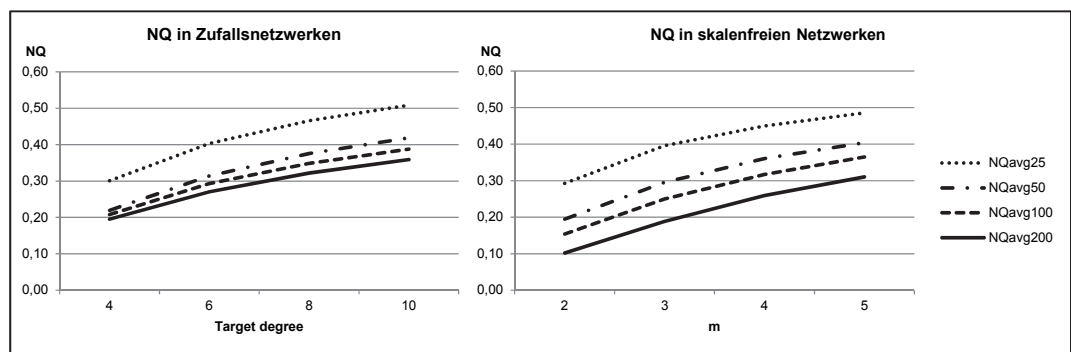


Abb. 64: NQ unterschiedlicher Netzwerke in Abhängigkeit von der Gradanzahl  
(Quelle: Eigene Darstellung)

Für beide Netztypen lässt sich eine deutliche Zunahme der NQ bei Erhöhung der Gradanzahl, d. h. der Dichte der Netze erkennen. Skalenfreie Netzwerke weisen dabei jedoch stärkere Unterschiede der einzelnen Größenklassen und – insbesondere bei großen Netzwerken – eine erheblich schlechtere NQ auf. Die generelle Steigerung der NQ bei steigender Dichte entspricht dem in Abschnitt 5.4 postulierten naheliegenden Zusammenhang. Je mehr Verbindungen ein Peer hat, desto größer ist im Mittel die Wahrscheinlichkeit, „gute“ Routings zu bilden, die zum einen Verbindungen mit hoher MQ erhalten und zum anderen eine geringere durchschnittliche Pfadlänge aufweisen und damit die NQ steigern bzw. weniger

<sup>801</sup> Vgl. die Hinweise in Tab. 14. Die Zufallsnetzwerke mit  $d \in [5; 7; 9]$  besitzen kein direktes Äquivalent. Ihre Verwendung würde die Zufallsnetzwerke überrepräsentieren.

stark verringern. Mit einer analogen Begründung ist das bessere Abschneiden kleiner Netzwerke zu erklären. Dieser Zusammenhang lässt sich auch bei einer Auswertung des dritt-einflussreichsten Faktors auf die NQ, der PMQG, nachweisen. Abb. 65 zeigt für Zufalls- und skalenfreie Netzwerke getrennt den Verlauf der NQ bei steigender PMQG.

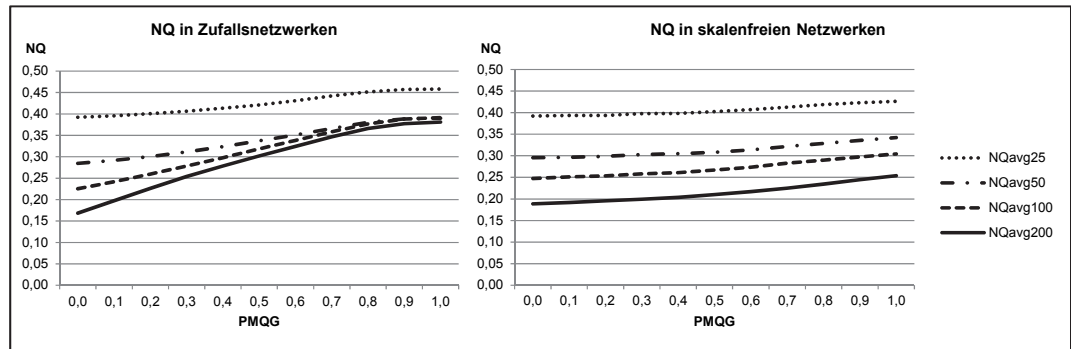


Abb. 65: NQ unterschiedlicher Netzwerke in Abhängigkeit von der PMQG

(Quelle: Eigene Darstellung)

Beide Netzwerktypen weisen zwangsläufig einen deutlichen, monotonen Anstieg der NQ bei Erhöhung der PMQG auf, da eine präzisere Vorhersage der MQ immer zu einer mindestens gleichguten Wahl führen muss wie die zufällige Auswahl – unabhängig von der definierten ‚Richtung‘ der MQ-Prüfung.<sup>802</sup> Dieser Anstieg ist in Zufallsnetzwerken jedoch erheblich größer als in skalenfreien Netzwerken. Zudem nähern sich die NQ-Werte der (mittel-)großen Netzwerke ( $n = 50; 100; 200$ ) für PMQG-Werte ab etwa 0,70 deutlich an, was einen erheblichen Einfluss der PMQG kennzeichnet. Der monoton steigende, fast lineare Verlauf der Kurven erlaubt zudem für weitere Analysen die Reduktion auf einen kleineren Wertebereich der PMQG ohne Verlust der allgemeinen Aussagekraft. So bilden extreme Werte nahe null keinen realen Fall ab, da selbst bei kürzester Betrachtung zweier Schemata grundlegende Aussagen über die Überführbarkeit getroffen werden können. Auch führen entsprechende automatisierte Algorithmen ohne weiteres Vorwissen zu statistisch signifikanten Korrelationen mit manuellen Überführungen mit  $\rho \geq 0,4$ . Bei entsprechender Voranalyse oder menschlicher Hilfe er-

<sup>802</sup> Bei einem ‚selbstlosen‘ Vorgehen (‚selfish‘-Parameter: ‚false‘) wird die mögliche maximale MQ der Partner-Peers erhöht, was bei nachgelagerter oder erneuter Routing-Bestimmung wiederum die RQ erhöht.



hört sich dieser Wert auf  $\rho \cong 0,7$ .<sup>803</sup> Hingegen erscheint die (fast) perfekte Vorhersage der MQ unrealistisch, da in diesem Falle bereits (nahezu) vollständige Überführungen gebildet worden sein müssten. Vor diesem Hintergrund und unter Berücksichtigung der starken Annäherung der NQ (bei Zufallsnetzwerken) für PMQG-Werte von 0,8 bis 1,0 wird für die weiteren Simulationen nur der PMQG-Wertebereich  $[0,4; 0,8]$  verwendet, was die Anzahl der durchzuführenden Simulationen und damit auch die Verzerrung durch ‚unrealistische‘ Parameterbelegungen verringert.<sup>804</sup>

Die höhere NQ von Zufallsnetzwerken zeigt sich durchgängig bei allen Netzwerkgrößen (vgl. Abb. 66 (links)). Sie erreichen diese trotz höherer durchschnittlicher Pfadlänge (vgl. Abb. 66 (rechts)), welche eigentlich – so wie eine allgemein höhere Ausfallrobustheit – ein distinktives Merkmal skalenfreier Netzwerke darstellt.<sup>805</sup> Allerdings weisen in der durchgeführten Simulation die skalenfreien Netzwerke auch in Bezug auf den für die Robustheit abprüfbaren Clustering-Koeffizienten<sup>806</sup> deutlich schlechtere Werte auf (vgl. Abb. 66 (rechts)).

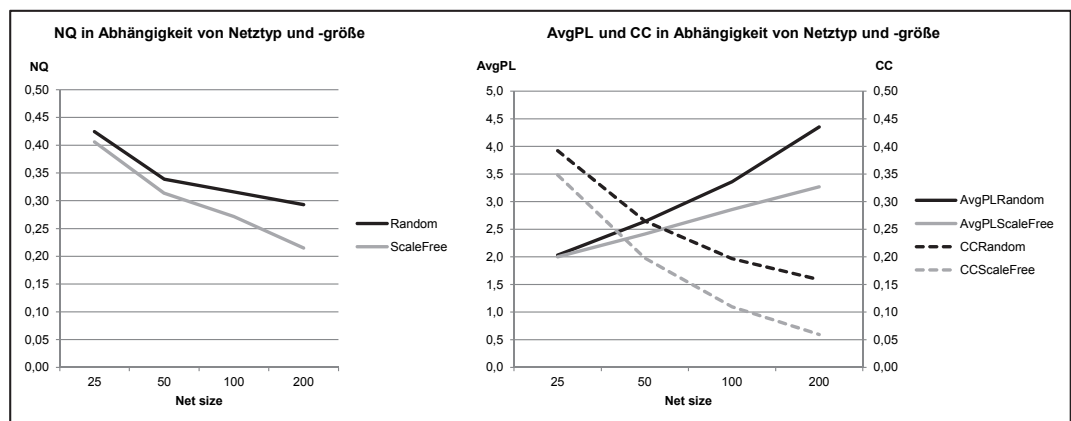


Abb. 66: NQ, AvgPL und CC in Abhängigkeit von Netztyp und -größe  
(Quelle: Eigene Darstellung)

<sup>803</sup> Vgl. Abschnitt 4.3.3. Die Werte geben die Erkenntnisse zu der automatisierten Überführung von Ontologie-Elementen unter Verwendung unterschiedlicher Algorithmen und im Vergleich zu rein manuellen Überführungen wieder (vgl. Bernstein et al. (2005): 1360f.).

<sup>804</sup> Für umfangreiche Auswertungen ist die Reduzierung der Simulationen auch aus pragmatischen Gründen wünschenswert, da durch diese Maßnahme auch die notwendige Rechenzeit zur Erstellung der Simulationen verringert wird. Zur Einordnung der Zeiten lässt sich das zur Verfügung stehende Testsystem (Intel-Core5-CPU mit 2,5 GHz, 8 Gigabyte Arbeitsspeicher) angeben: Es führt 1.000 Netzwerksimulationen ( $n = 200$ ) ohne Berechnung der TNQ in ca. 20 Minuten, mit Berechnung der TNQ in ca. 60 Minuten durch.

<sup>805</sup> Vgl. die Ausführungen zu skalenfreien Netzwerken und deren Eigenschaften in Abschnitt 5.2.2.

<sup>806</sup> Vgl. Abschnitt 5.3.4.

Da die bisherigen Auswertungen Durchschnittswerte über die algorithmusspezifischen Parameter beschreiben, besteht die Möglichkeit, dass einzelne Parameter die Durchschnittswerte insbesondere der skalenfreien Netzwerke stark negativ beeinflussen. Abb. 67 stellt daher zum Abgleich den Einfluss der spezifischen Parameter ‚sorted‘ und  $\mu$  dar, die dazu geeignet sind, bei höheren Werten eine stärker PMQG-orientierte Nachbarschaftswahl zu verursachen. Ein relevanter Einfluss der Parameter ist jedoch für keine der Kombinationen zu erkennen.<sup>807</sup>

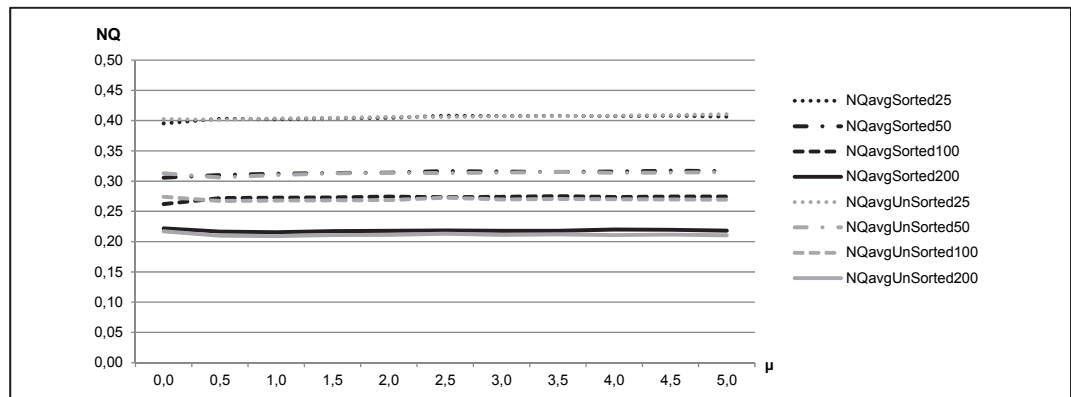


Abb. 67: NQ skalenfreier Netzwerke in Abhängigkeit von  $\mu$  und Knotensortierung (Quelle: Eigene Darstellung)

Abschließend zeigt Abb. 68 den Einfluss des letzten bisher ungeprüften Parameters ‚selfish‘ auf die NQ für Zufalls- und skalenfreie Netzwerke. Auch dieser weist für die skalenfreien Netzwerke nur kleine Unterschiede aus.

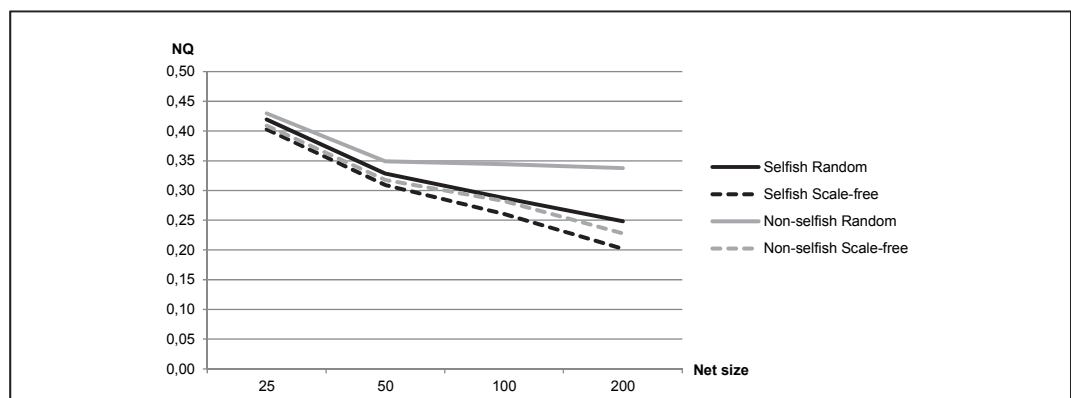


Abb. 68: NQ in Abhängigkeit von Netztyp, Netzgröße und ‚selfish‘-Parameter (Quelle: Eigene Darstellung)

<sup>807</sup> Der minimale Einfluss einzelner Parameter lässt sich auch für Zufallsnetzwerke bei Betrachtung der ‚core net size‘ erkennen, die bei keiner Netzgröße eine Wertabweichung von mehr als 0,32% verursacht. Vgl. für Detailanalysen von Zufallsnetzwerken und deren Parameter Abschnitt 6.4.3.3.

Auffällig ist allerdings die konstant hohe NQ für Zufallsnetzwerke bei ‚selbstloser‘ Nachbarschaftsauswahl. Sie zeigt sich ganz besonders deutlich bei hohen PMQG-Werten (vgl. Abb. 69).

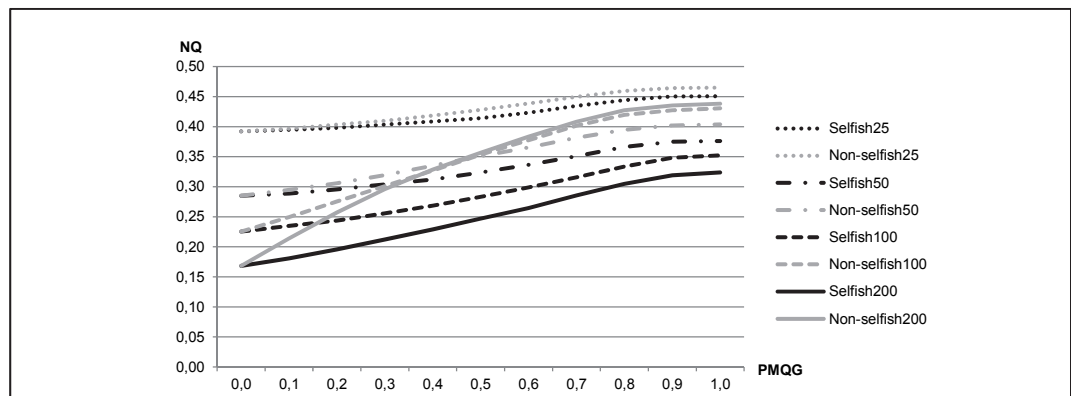


Abb. 69: NQ von Zufallsnetzwerken in Abhängigkeit von PMQG und ‚selfish‘-Parameter

(Quelle: Eigene Darstellung)

Der Effekt entsteht, da bei hohen PMQG-Werten die Nachbarschaftswahl auf (annähernd) den MQ-Werten basiert, die nach Erzeugung des Netzes zur Errechnung der NQ tatsächlich herangezogen werden.<sup>808</sup> Dabei wählen neu eingefügte Knoten diejenigen Knoten als Nachbar aus, denen sie besonders viele Informationen bereitstellen können, was ganz besonders auf Knoten mit sehr geringem Schemaumfang zutrifft. Für diese entsteht dadurch ein hochangepasstes Netzwerk, in dem MQ-Werte nahe eins erreicht werden können, was die durchschnittliche Netzqualität stark erhöht, obwohl die tatsächlich übertragene Datenmenge – bezogen auf das ganze Netz – möglicherweise zurückgeht, da die neu hinzugefügten Knoten ggf. ‚schlechte‘ Verbindungen aus eigener Rezipientensicht erzeugen. Der Effekt wurde bereits bei der Vorstellung der Einflussfaktoren auf die Netzwerküte diskutiert und zeigt sich hier am Modell der Simulation.<sup>809</sup> Es ist daher notwendig für die Einordnung der gezeigten Ergebnisse, die relevanten Parameter an der TNQ zu prüfen, um sicherzustellen, dass die ausführlich dargelegte Überlegenheit von Zufallsnetzwerken gegenüber skalenfreien Netzwerken auch in Hinblick auf die tatsächlich übertragenen Daten existiert, und um den Einfluss der ‚selbstsüchtigen‘ oder ‚selbstlosen‘ Nachbarbarschaftswahl korrekt zu beurteilen.

<sup>808</sup> Vgl. Abschnitt 6.3.2.2.

<sup>809</sup> Vgl. Abschnitt 5.4.

### 6.4.3.2 Ergebniseinordnung auf Basis der tatsächlichen Netz-Qualität

Zum Vergleich von NQ und TNQ wurde eine zweite Simulationsreihe erstellt, bei der 19.200 Netze generiert wurden. Auf Basis der vorgestellten Erkenntnisse wurden dafür die überprüften Parameterbereiche in Teilen eingeschränkt, im Gegenzug allerdings die Anzahl der Seeds erhöht. Tab. 20 gibt eine Übersicht der verwendeten Parameterkonfigurationen.

Parameter	Wertebereich
Basic algorithm	Random; Scale-free
Net size	25;50;100;200
Target degree	[4;10] @ 2
m	[2;5] @ 1
Core net size	1
Seed	[1;6] @ 1
PMQG	[0,4;0,8] @ 0,1
Target degree spread	[0,1;1,0] @ 0,1 <sup>810</sup>
Sorted node list	True
$\mu$	[0,00; 2,25] @ 0,25 <sup>811</sup>
Selfish approach	True; False
MQ ( $\alpha$ , $\beta$ , $\gamma$ )	(0,60; 0,05; 0,35)

Tab. 20: Parameterkonfigurationen der zweiten Netzwerk-Simulationsreihe  
(Quelle: Eigene Darstellung)

Die Ergebnisse bestätigen hinsichtlich der NQ die Werte der ersten Simulationsreihe für die bereits vorgestellten Parameter. Sie validieren zudem die Erkenntnisse der Betrachtung von NQ und TNQ, die Rahmen der Ermittlung geeigneter t-Normen erlangt wurden, insbesondere im Hinblick auf die netzgrößenabhängige NQ-/TNQ-Verteilung und die Aufteilung der Datenpunkte in mehrere Bereiche, die eine unterschiedlich starke Steigung von Regressionsgeraden aufweisen.<sup>812</sup> Abb. 70 stellt für die unterschiedlichen Konfigurationen die TNQ in Abhängigkeit von der NQ dar. Die Ergebnisse sind zudem in Tab. 21 zusammengefasst.

<sup>810</sup> Ein Wert von 0,0 verhindert, dass ein neu eintretender Knoten eine MQ-basierte Nachbarschaftswahl vornimmt, da ihm nur die letzten in das Netz eingetretenen Teilnehmer zur Verfügung stehen. Ein solches Netzwerk ist nicht in der Lage, Untersuchungen zur Nachbarschaftswahl, die Gegenstand der folgenden Abschnitte sind, zu unterstützen. Der Wert wurde für die übergreifende Analyse in Abschnitt 6.4.3.1 der Vollständigkeit halber abgebildet, wird jedoch – wie auch einzelne Werte der PMQG – für die weiteren Analysen nicht berücksichtigt.

<sup>811</sup> Der minimale Einfluss von  $\mu$  auf die Netzwerkgröße wurde bereits in Abschnitt 6.4.3.1 gezeigt. Die  $\mu$ -Werte wurden hier so gewählt, dass eine gleiche Anzahl von Zufallsnetzwerken und skalenfreien Netzwerken generiert wird, um eine Vergleichbarkeit der Ergebnisse zu gewährleisten und dabei keine Netzwerkkonfigurationen mehrfach zu verwenden.

<sup>812</sup> Vgl. Abschnitt 6.4.2, insbesondere Abb. 60.

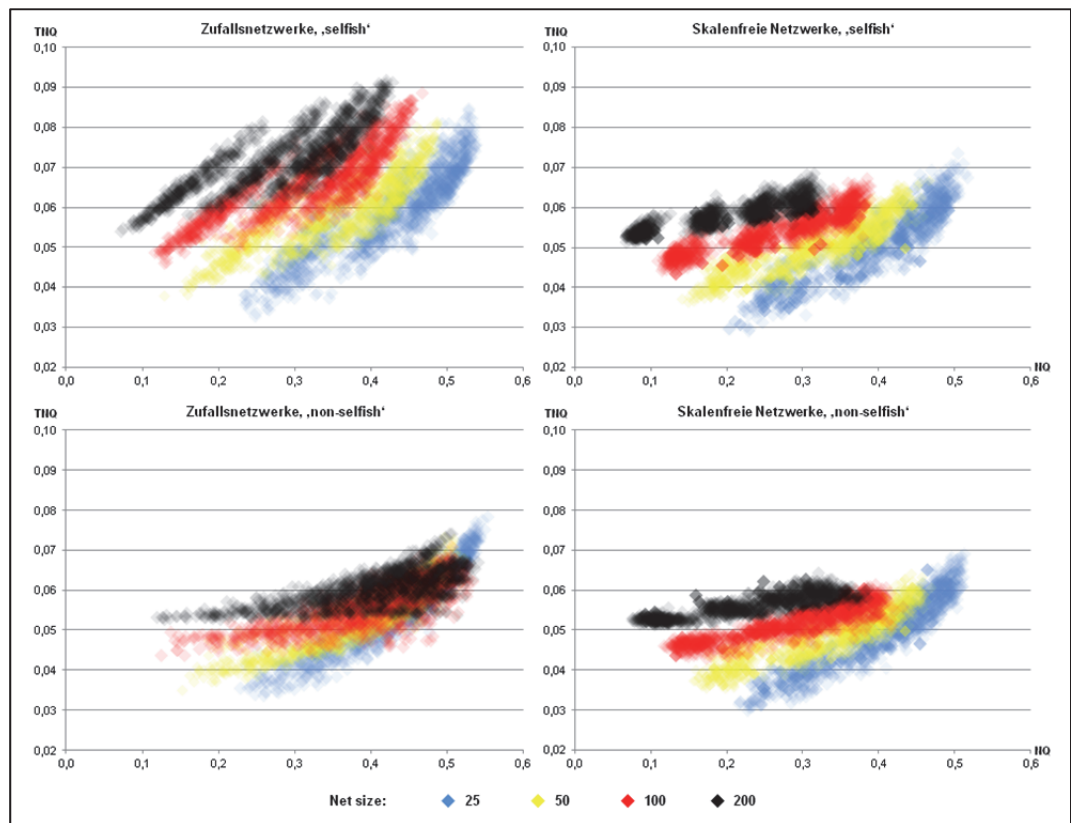


Abb. 70: TNQ in Abhängigkeit von der NQ für unterschiedliche Netzwerke  
(Quelle: Eigene Darstellung)

Netzgröße	Netztyp	,selfish'	NQ	TNQ
25	Zufallsnetzwerk	true	0,4282	0,0592
		false	0,4426	0,0559
	Skalenfreies Nw.	true	0,3958	0,0507
		false	0,4042	0,0494
50	Zufallsnetzwerk	true	0,3547	0,0589
		false	0,3890	0,0529
	Skalenfreies Nw.	true	0,3221	0,0505
		false	0,3340	0,0485
100	Zufallsnetzwerk	true	0,3161	0,0659
		false	0,3957	0,0558
	Skalenfreies Nw.	true	0,2639	0,0539
		false	0,2877	0,0516
200	Zufallsnetzwerk	true	0,2816	0,0722
		false	0,4033	0,0602
	Skalenfreies Nw.	true	0,2031	0,0581
		false	0,2334	0,0561

Tab. 21: Durchschnittliche NQ und TNQ für unterschiedliche Netzwerke  
(Quelle: Eigene Darstellung)

Die gezeigten Konfigurationen sind auf identische Achsenabschnitte skaliert, sodass sich unmittelbar erhebliche Unterschiede in der NQ und TNQ erkennen lassen. Die auffällige ‚Streifen‘- oder ‚Gruppen‘-Bildung bei Netzwerken mit ‚selfish‘-Konfiguration spiegelt dabei die unterschiedlichen durchschnittlichen Gradanzahlen wider, deren starker Einfluss (innerhalb der Netzgrößen) bereits bei der reinen NQ-Betrachtung gezeigt werden konnte.<sup>813</sup> Als Referenzgröße für die Einordnung der TNQ dienen vollständig verbundene Netze, die unter den gegebenen DWH-Strukturen der einzelnen Datenknoten die maximal mögliche Übertragbarkeit von Daten gewährleisten. Die TNQ ist auch hier Netzgrößen-abhängig und steigt mit der Netzgröße auf einen Maximalwert von 0,1171 bei  $n = 200$ .<sup>814</sup>

Damit lassen sich sowohl grafisch als auch tabellarisch die folgenden Hypothesen ableiten:

- (H1) Netzwerke mit ‚selfish‘-Konfiguration erzielen für beide Netzwerktypen im Durchschnitt höhere Werte der TNQ als solche mit ‚non-selfish‘-Konfiguration.<sup>815</sup>
- (H2) Zufallsnetzwerke mit ‚selfish‘-Konfiguration erzielen zudem höhere Werte der TNQ als skalenfreie Netzwerke gleicher Konfiguration.
- (H3) Die TNQ steigt mit der Netzwerkgröße.
- (H4) Die TNQ steigt mit der Netzwerkgröße für Netzwerke mit  $n \geq 50$ .

Dabei wird (H4) gesondert zu (H3) betrachtet, da kleine Netzwerke (hier:  $n = 25$ ) bereits bei der Betrachtung der NQ deutliche Unterschiede zu größeren Netzwerken aufweisen, was z. B. mit der Wahl der Gradanzahlen zu begründen ist, da in kleinen Netzwerken auch niedrige Gradanzahlen zu hoher Dichte führen.

<sup>813</sup> Vgl. Abschnitt 6.4.3.1. Der Einfluss der Gradanzahlen auf die Interpretation der Ergebnisse hinsichtlich des Zusammenhangs von NQ und TNQ wird in Abschnitt 6.4.4 ausführlicher diskutiert.

<sup>814</sup> Errechnet wurden auch hier sechs unterschiedliche Werte mit abweichender Seed-Konfiguration zur Sicherstellung unterschiedlicher Knotenaufbauten. Angegeben ist, um jede Verzerrung zu vermeiden, der Maximalwert anstelle des Durchschnittswerts. Die Ergebnisse zeigen, dass die absoluten TNQ-Werte der Netzwerk-Simulationen zwar – gemessen am theoretischen Maximalwert von 1,0 – vergleichsweise niedrig sind, dies jedoch vorrangig durch die mangelnde Überführbarkeit begründet ist und nicht durch einen Mangel an direkten Verbindungen. Für die übrigen Netzgrößen gelten zudem noch geringere Maximalwerte ( $n = 25: 0,0951; n = 50: 0,1003; n = 100: 0,1111$ ).

<sup>815</sup> Geprüft wird hier vor allem die TNQ, da die NQ zwar ein gutes Abbild der TNQ darstellen soll (was durch die Wahl der t-Norm (vgl. Abschnitt 6.4.2) und eine korrekte Zuordnungsfunktion (vgl. Abschnitt 6.4.4) sichergestellt wird), die TNQ aber die maßgebliche Größe bei der Auswahl der Grundkonfiguration darstellt, da sie die tatsächliche Güte des Netzwerks abbildet.

Zur Überprüfung der Hypothesen wurden entsprechende Null- und Alternativhypothesen in Bezug auf die Durchschnittswerte gebildet und die erkannte Abweichung der Werte auf statistische Signifikanz geprüft. Als Standard-Test ist hierfür der *t-Test* etabliert.<sup>816</sup> Dieser setzt eine Normalverteilung der überprüften Werte voraus, die hier nicht gegeben ist.<sup>817</sup> Der t-Test verhält sich allerdings robust bei Verletzung der Voraussetzung, sofern die Stichprobengrößen identisch sind (was hier durch die Parameterbelegung der zweiten Simulationsreihe gesichert wurde) und die Verteilung der zugrunde liegenden Werte zumindest eingipflig-symmetrisch ist.<sup>818</sup> Für alle vorgestellten Stichprobenumfänge wurde diese Eigenschaft mit Hilfe einer Histogramm-basierten Betrachtung überprüft und positiv bestätigt, sodass der t-Test anstelle komplexerer Verfahren verwendet werden kann.<sup>819,820</sup> Alle vorgestellten Untersuchungen wurden dabei auf einem Signifikanzniveau von  $\alpha = 0,05$  durchgeführt, was als übliche Größe angesehen wird.<sup>821</sup> Tab. 22 stellt die erzielten p-Werte aller durchgeführten Tests dar.

---

<sup>816</sup> Vgl. Wooldridge (2013): 113–115. Der Test leitet seinen Namen aus der studentschen t-Verteilung ab, der Mittelwerte von Stichproben bei unbekannter Varianz der Grundgesamtheit folgen.

<sup>817</sup> Vgl. die vorherige Untersuchung im Rahmen der t-Norm-Ermittlung in Abschnitt 6.4.2.

<sup>818</sup> Vgl. Bortz/Schuster (2010): 120f.; sowie die dort angegebenen Quellen.

<sup>819</sup> Als alternatives Standard-Verfahren zur Überprüfung bei Verletzung der Bedingungen dient der Wilcoxon-Mann-Whitney-Test (auch: ‚U-Test‘), der auch unterschiedliche Stichprobenumfänge zulässt und die Gleichartigkeit der Verteilungen in den Stichproben prüft (vgl. Neuhäuser (2014): 1656f.).

<sup>820</sup> Bei der Durchführung der t-Tests wurde zudem berücksichtigt, dass die Varianzen der Stichproben unterschiedlich sein können. Für alle Tests wurden die Stichproben mit einem F-Test dahingehend überprüft (vgl. zum Vorgehen Bortz/Schuster (2010): 128f.) und die Errechnung der t-Statistik entsprechend durchgeführt.

<sup>821</sup> Vgl. Backhaus (2011): 78; Bortz/Schuster (2010): 11; Wooldridge (2013): 115f. Die Prüfgröße gegen das Signifikanzniveau  $\alpha$  ist der *p-Wert* (Signifikanzwert).

Stichprobe 1			Stichprobe 2			p-Wert	Pro-Indikator für	Contra-Indikator für
Größe	Typ	'selfish'	Größe	Typ	'selfish'			
25	Zufallsnetzwerk	true	25	Zufallsnetzwerk	false	0,0000	(H1)	
		true			true	0,0000	(H2)	
	Skalenfreies Nw.	true		Skalenfreies Nw.	false	0,0003	(H1)	
50	Zufallsnetzwerk	true	50	Zufallsnetzwerk	false	0,0000	(H1)	
		true			true	0,0000	(H2)	
	Skalenfreies Nw.	true		Skalenfreies Nw.	false	0,0000	(H1)	
100	Zufallsnetzwerk	true	100	Zufallsnetzwerk	false	0,0000	(H1)	
		true			true	0,0000	(H2)	
	Skalenfreies Nw.	true		Skalenfreies Nw.	false	0,0000	(H1)	
200	Zufallsnetzwerk	true	200	Zufallsnetzwerk	false	0,0000	(H1)	
		true			true	0,0000	(H2)	
	Skalenfreies Nw.	true		Skalenfreies Nw.	false	0,0000	(H1)	
50	Zufallsnetzwerk	true	25	Zufallsnetzwerk	true	0,2339		(H3)
100			50			0,0000	(H3), (H4)	
200			100			0,0000	(H3), (H4)	

Tab. 22: *p*-Werte der *t*-Tests zur Überprüfung einer Abweichung der durchschnittlichen TNQ bei unterschiedlicher Konfiguration

(Quelle: Eigene Darstellung)

Die durchgeführten Untersuchungen zeigen, dass Zufallsnetzwerke mit ‚selfish‘-Konfiguration sowohl skalenfreie Netzwerke (unabhängig von dem dort gewählten ‚selfish‘-Parameter) als auch Zufallsnetzwerke mit ‚non-selfish‘-Konfiguration dominieren. Die Hypothesen (H1) und (H2) werden daher unter Berücksichtigung des gegebenen Signifikanzniveaus als zutreffend angenommen.<sup>822</sup>

Es zeigt sich zudem, dass eine Erhöhung der Netzgröße die TNQ signifikant erhöht, sofern die Teilnehmerzahl eine bestimmte Größe überschreitet (hier:  $n \geq 50$ ). Dies entspricht einer Bestätigung von (H4) und einer Ablehnung von (H3). Für kleine Netzwerke lässt sich dieser Effekt durch den beschriebenen hohen Einfluss der Gradanzahl erklären, die hier nur im Durchschnitt betrachtet wurde. Er führt dazu, dass bei der Vergrößerung der Netzwerke die Dichte stark zurückgeht und NQ und TNQ zunächst sinken.<sup>823</sup> Dieser ‚Nachteil‘ bei der Erstellung größe-

<sup>822</sup> Tatsächlich zeigen die *p*-Werte, dass auch höhere Signifikanzniveaus, d. h. kleinere  $\alpha$ -Werte, erreicht werden können.

<sup>823</sup> Der Effekt ist mit der gleichen Methodik, die für die bisherigen Überprüfungen angewendet wurde, auch statistisch für die Betrachtung hoher Gradanzahlen ( $d = 10$ ) nachweisbar, er trifft allerdings nicht mehr auf kleine Gradanzahlen ( $d = 4$ ) zu ( $\alpha = 0,05$ ). Dies ist plausibel, da hier nur kleine Netzwerke betrachtet wurden und sich die durch die Gradanzahl positiv beeinflusste Dichte anti-proportional zur Teilnehmerzahl ermittelt (vgl. Abschnitt 5.2.2).



rer Netze wird allerdings ab einer bestimmten Teilnehmerzahl überkompensiert durch die steigende Auswahl an Partnern, die zunehmend individuell bessere Verbindungen und Routings ermöglichen. Das Netzwerk erreicht ab diesem Punkt eine ‚kritische Masse‘, die in den Wirtschafts- und Sozialwissenschaften bereits seit Längerem für diverse Netzwerkstrukturen und Einsatzfelder diskutiert und nachgewiesen wurde und teilweise eine Notwendigkeit für wirtschaftlich erfolgreiches Handeln darstellt.<sup>824</sup>

Auf Basis der Ergebnisse werden skalenfreie Netze nicht weiter betrachtet, da sie in keiner Konfiguration einen Vorteil gegenüber Zufallsnetzwerken aufweisen. Ebenfalls nicht weiter betrachtet werden Netzwerke mit ‚non-selfish‘-Konfiguration, da diese eine erheblich schlechtere TNQ als die Netzwerke mit ‚selfish‘-Konfiguration aufweisen und zudem eine starke Streuung der NQ bei schwach variierender TNQ aufweisen (vgl. Abb. 70), was eine Prognose der TNQ basierend auf der NQ erschwert und zu Fehlinterpretationen bei der Auswertung der NQ führen kann.<sup>825</sup>

#### 6.4.3.3 Detailanalyse relevanter Parameter

Die reine Betrachtung von Zufallsnetzwerken zeigt für die bekannten Größen eine ähnliche Verteilung der NQ-Werte wie bei der Gesamtbetrachtung.<sup>826</sup> Dabei werden allerdings keine Netzwerke mit äußerst schlechter NQ nahe null erzeugt. Die Werte der NQ steigen zudem in Bezug auf die untersuchten Größenklassen bei Erhöhung der Gradanzahl und PMQG ähnlich stark (vgl. Abb. 71) sowie in beiden Fällen monoton. Auffällig ist, dass für kleine Netzwerke ( $n = 25$ ) der Unterschied zur nächst größeren Netzwerkklassse in Bezug auf die NQ größer ist, als zwischen den übrigen Klassen ( $n \in \{50; 100; 200\}$ ). Dies bestätigt die bereits herausgearbeitete Sonderposition kleiner Netzwerke in Bezug auf die überprüften Qualitäten.<sup>827</sup>

<sup>824</sup> Vgl. dazu ausführlich die Betrachtung von Grenzkosten und -nutzen von Netzwerkteilnehmern (u. a. in Telekommunikationsnetzwerken) bei KATZ/SHAPIRO (vgl. Katz/Shapiro (1994): 93–97). Ältere Untersuchungen im gleichen Feld finden sich bei OREN/SMITH, die auch den Begriff der ‚kritischen Masse‘ prägen (vgl. Oren/Smith (1981): 467, 472–474). Untersuchungen mit dem Fokus auf gemeinschaftliches Handeln (z. B. im politisch motivierten Umfeld) basieren maßgeblich auf MARWELL/OLIVER/PRAHL (vgl. Marwell/Oliver/Prahl (1988): 502–534) und lassen sich auf moderne soziale Medien übertragen (vgl. Ghobadi/Clegg (2015): 52–56).

<sup>825</sup> Vgl. Abschnitt 6.4.3.1.

<sup>826</sup> Vgl. Abschnitt 6.4.3.1.

<sup>827</sup> Vgl. Abschnitt 6.4.3.2.

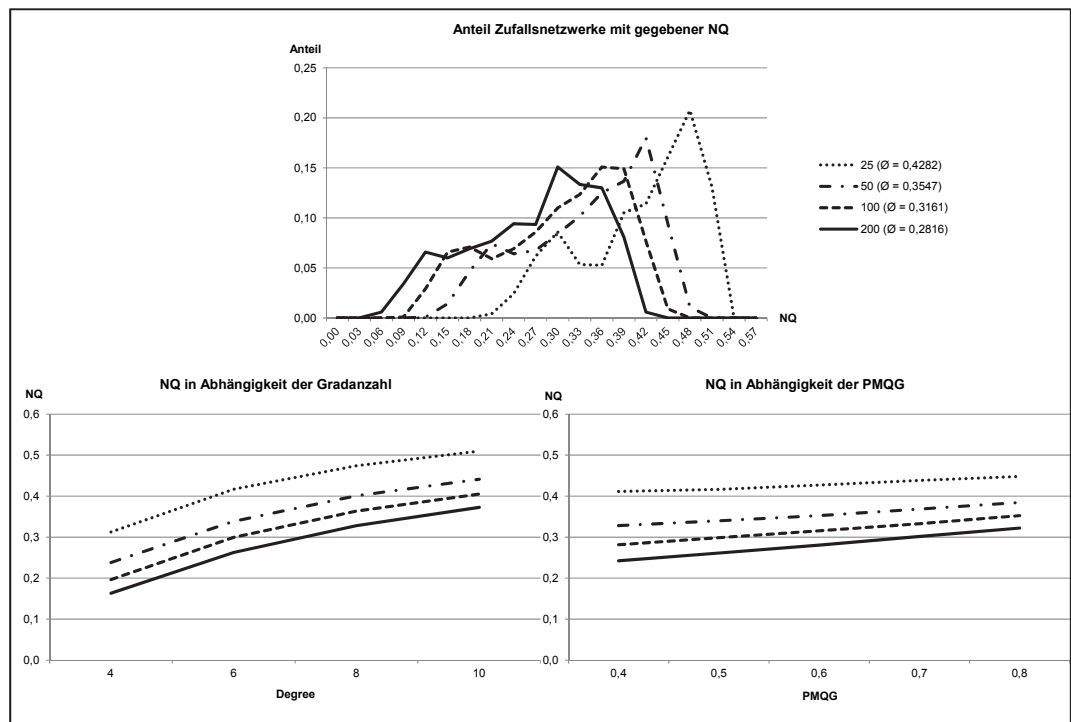


Abb. 71: Netzwerkanteile und NQ-Werte für Zufallsnetzwerke  
(Quelle: Eigene Darstellung)

Als letzter nicht betrachteter Parameter kann der für Zufallsnetzwerke spezifische ‚Target degree spread‘ (kurz: ‚Spread‘) betrachtet werden, der angibt, in welchem Maße eine Abweichung je Peer von der Zielgradanzahl ‚Degree‘ erlaubt ist. Abb. 72 stellt den Verlauf der NQ in Abhängigkeit vom Spread dar.

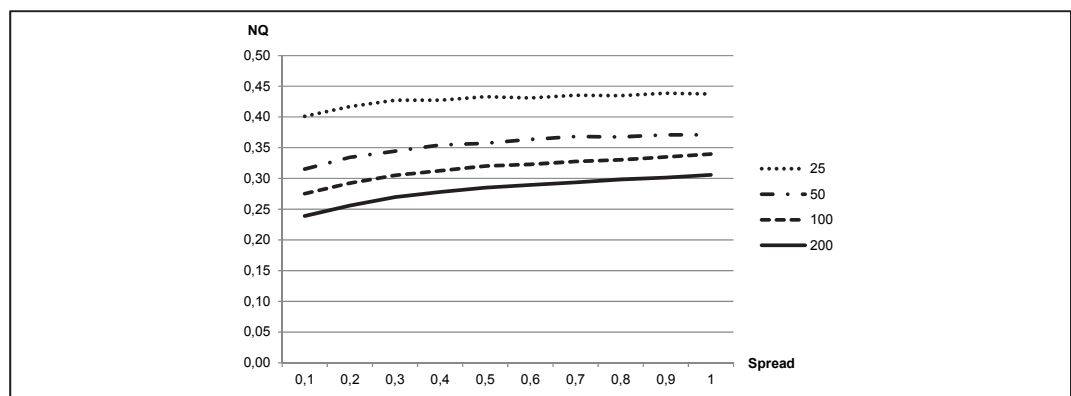


Abb. 72: NQ von Zufallsnetzwerken in Abhängigkeit vom Spread  
(Quelle: Eigene Darstellung)

Der Spread weist einen positiven, fast monoton steigenden Einfluss auf die NQ auf. Dabei zeigt sich ein etwas stärkerer Einfluss bei der Erhöhung sehr kleiner Werte. Diese entsprechen einer Aufrechterhaltung der Zielgradanzahl in den Netzen, die die Variabilität der Knoten stark einschränkt und z. B. verhindert, dass

Knoten mit einem sehr detaillierten Schema stärker verknüpft werden, oder Knoten mit sehr ähnlichen Schemata eine Verbindung schaffen können, sobald bereits einer der Knoten seine Maximalanzahl an Verbindungen erreicht hat. Eine starke Erhöhung des Parameters führt im Übrigen nicht zu einer Imbalance des Netzes im Sinne der RQ einzelner Peers. Zum Nachweis wurde die *Standardabweichung* (SD) der RQ, d. h. der Peer-spezifischen Bestandteile der NQ, je Netz ermittelt. Ein Vergleich der durchschnittlichen SD zeigt, dass auch bei sehr großer Variabilität in der Gradanzahl zwischen einzelnen Peers eine unveränderte Gleichmäßigkeit der RQ für die Netzteilnehmer zu erwarten ist (vgl. Abb. 73).

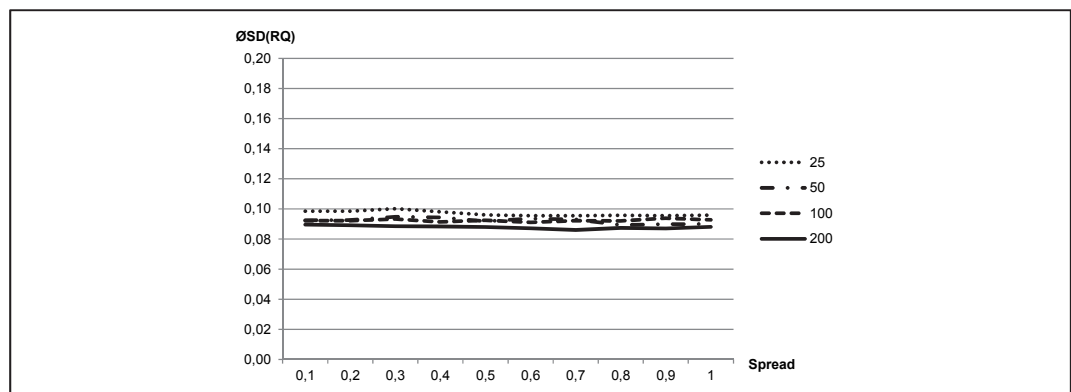


Abb. 73: Durchschnittliche Standardabweichung der RQ in Zufallsnetzwerken in Abhängigkeit vom Spread

(Quelle: Eigene Darstellung)

Alle getesteten Parameter (Netzgröße, Gradanzahl, PMQG und Spread) weisen somit einen erkennbaren Einfluss auf die NQ (und damit die TNQ) auf. Er ist fast durchgängig monoton und positiv und weist nur für die Netzgröße in Bezug auf die NQ einen negativen Zusammenhang auf, was durch die Berechnung der NQ und die steigenden Pfadlängen in großen Netzwerken zu erklären ist. Für die praktische Anwendung der Netzwerke muss abschließend geklärt werden, ob die Parameter unabhängig voneinander beeinflussbar sind oder ob Wechselwirkungen auftreten, die Erhöhungen von Einflussgrößen verstärken oder abschwächen. Zudem ist die Stärke des Einflusses entscheidend, da Netzteilnehmern in der Regel nur begrenzte Ressourcen zur Verfügung stehen, die möglichst effektiv eingesetzt werden sollen. Abb. 74 zeigt dazu exemplarisch für die vier betrachteten Netzgrößen den Einfluss von PMQG und Spread im Vergleich.

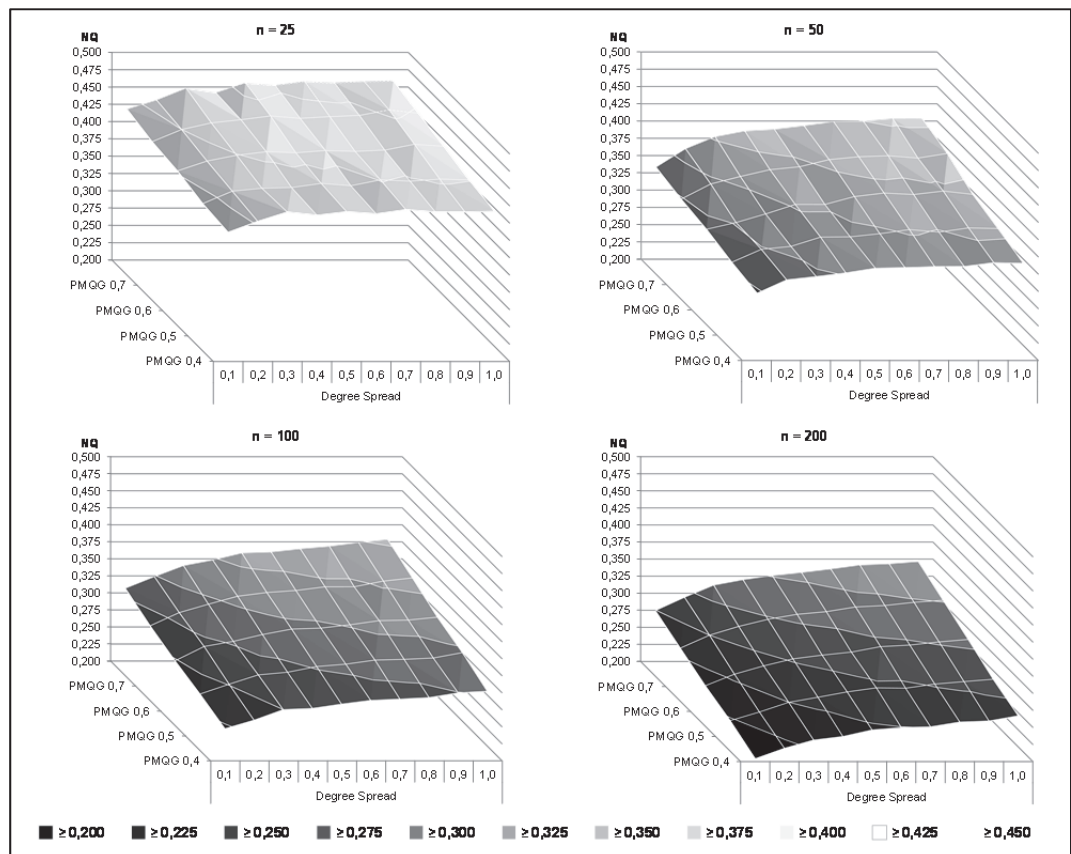


Abb. 74:  $NQ$  in Abhängigkeit von  $PMQG$ ,  $Spread$  und Netzwerkgröße  
(Quelle: Eigene Darstellung)

Deutlich erkennbar ist der im Mittel erheblich größere Einfluss der Netzgröße auf die  $NQ$  im Vergleich zu  $PMQG$  und  $Spread$ , während für diese beiden Größen zunächst kein dominanter Faktor im geprüften Wertebereich erkennbar ist. Es zeigt sich aber eine Abhängigkeit der Steigung der Kurve – insbesondere für kleine Werte von  $PMQG$  und  $Spread$  – von der Netzgröße. Der Einfluss einer  $PMQG$ - und  $Spread$ -Erhöhung ist für  $n = 200$  offenbar größer als für kleinere Netze. Zur Aufdeckung dieser Wechselwirkungen wurde eine multiple Regressionsanalyse auf den Werten der zweiten Simulationsreihe durchgeführt.<sup>828</sup> Die Anwendbarkeit dieser Analyse ist u. a. von der Annahme linearer Zusammenhänge der unabhängigen und abhängigen Variablen beeinflusst. Solche können auf Basis der vorgestellten Parameter-Auswertungen als näherungsweise vorhanden angenommen werden, insbesondere, da hier *Interaktionsterme* verwendet werden. Diese beschreiben eine Verknüpfung zweier oder mehrerer unabhängiger Variablen und werden üblicherweise durch die multiplikative Verknüpfung der Variablen gebil-

<sup>828</sup> Vgl. zum Einsatz der multiplen Regression Abschnitt 6.4.1.

det.<sup>829</sup> Es lassen sich auch komplexere Kombinationen einsetzen, deren Interpretation dann schwieriger wird, die aber ggf. bestimmte Sachverhalte zutreffender abbilden.<sup>830</sup> Für die vorliegenden Netzwerke ist dies z. B. durch die ‚Dichte‘ als Verknüpfung von Netzgröße und Gradanzahl (bzw. der dadurch bestimmten Verbindungsanzahl) gegeben. Die nachfolgende Analyse verzichtet jedoch explizit auf die Verwendung der Dichte, da zum einen die Interpretation des Einflusses in diesem Fall von derjenigen der anderen Interaktionsterme abweichen würde, zum anderen eine direkte Beeinflussbarkeit der NQ durch die Festlegung der Dichte impliziert würde. Diese ist jedoch nicht gegeben, da die Dichte nur eine errechnete Größe, nicht aber einen definierbaren Parameter im Rahmen der Simulationen beschreibt. Sie ist zudem – wie für einen Interaktionsterm der Netzgröße angenommen – nicht Netzgrößen-unabhängig, sondern beeinflusst die NQ für verschiedene Netzgrößen unterschiedlich stark (vgl. Abb. 75), was wiederum Fehlinterpretationen begünstigt.<sup>831</sup>

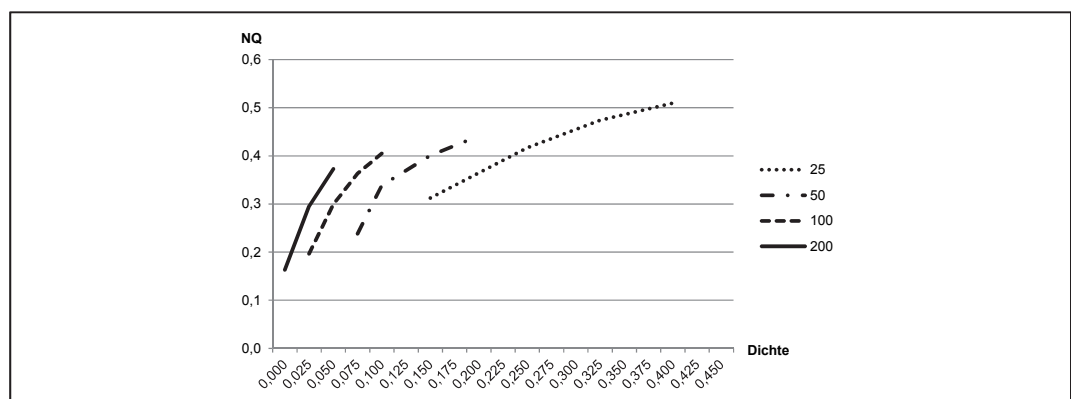


Abb. 75: NQ von Zufallsnetzwerken in Abhängigkeit von der Dichte

(Quelle: Eigene Darstellung)

Zur Durchführung der Regressionsanalyse wurden die Werte aller Variablen und Interaktionsterme standardisiert. Es ergeben sich daher für die Regression die

<sup>829</sup> Vgl. Backhaus (2011): 86–88. Der Einsatz der Interaktionsterme hebt in Teilen die Beschränkung der Linearität des Modells auf, da nur die einzelnen Terme miteinander linear verknüpft sein müssen, diese selbst aber auch komplexere Variablenkonstellationen abbilden können.

<sup>830</sup> Wooldridge (2013): 191.

<sup>831</sup> Vgl. für eine beispielhafte (Miss-)Interpretation von Interaktionstermen unter Berücksichtigung ‚interessanter‘ Kombinationen, d. h. Kombinationen, die eine bestimmte (Eigen-)Bedeutung aufweisen – wie hier die Dichte – Wooldridge (2013): 191f. Insbesondere, falls die konkreten Werte der Koeffizienten von Bedeutung für weitergehende Analysen sind, müssen die ermittelten Koeffizienten von Einzel- und Interaktionstermen zunächst auf ihre statistische Signifikanz geprüft werden. Für die hier vorliegenden Analysen soll nur die ungefähre Einflusshöhe bestimmt werden, sodass auf umfangreiche Tests (wie z. B. in Abschnitt 6.4.3.2 durchgeführt) verzichtet wird.

standardisierten Koeffizienten, die für die einzelnen Variablen vergleichbar sind und den unterschiedlichen Einfluss dieser ausdrücken.<sup>832</sup> Tab. 23 stellt die Ergebnisse der Analyse in Bezug auf die NQ nach absoluten Werten absteigend geordnet dar.

Parameter	Wert
Degree	0,9476
Net size	- 0,8186
Net size x PMQG	0,2339
PMQG	0,2028
Spread	0,1835
Degree x Spread	- 0,1777
Degree x PMQG	- 0,1577
Spread x PMQG	0,0874
Net size x Spread	0,0798
Net size x Degree	0,0615

Tab. 23: Einfluss von Variablen und Interaktionstermen auf die NQ<sup>833</sup>

(Quelle: Eigene Darstellung)

Der hohe Einfluss von Gradanzahl und Netzgröße entspricht den bisherigen Erkenntnissen und spiegelt auch die initial identifizierte Variabilität der NQ bei Veränderung dieser Werte wider.<sup>834</sup> Der entsprechende Interaktionsterm ist dabei der kleinste Wert der Reihe, sodass eine reine Erhöhung der Gradanzahl (nahezu) ohne Berücksichtigung der vorliegenden Netzgröße zu einer deutlichen Erhöhung der NQ führt. Wesentlich stärkere Interaktionen ergeben sich für die PMQG. Eine Erhöhung dieser beeinflusst die NQ insbesondere in großen Netzen positiv, hat jedoch bei einer hohen Gradanzahl einen verringerten Einfluss.

Für die vorliegenden Analysen ist allerdings zu berücksichtigen, dass die NQ die TNQ zwar grundsätzlich gut abbildet, die Prognosekraft jedoch erheblich von der Betrachtung einzelner Netzgrößen abhängt. Da die Netzgröße selbst ein Einflussfaktor der durchgeführten Regressionsanalyse ist, entfällt diese Anpassung hier, was die Aussagekraft der Analyse – insbesondere hinsichtlich der Effekte auf die TNQ – einschränkt. Tab. 24 zeigt daher die Ergebnisse einer analog durchgeführ-

<sup>832</sup> Vgl. Wooldridge (2013): 181f.

<sup>833</sup> Für die Analyse gilt  $R^2 = 0,8921$ .

<sup>834</sup> Vgl. Abschnitt 6.4.3.1

ten Regressionsanalyse mit Bezug auf die TNQ anstelle der NQ. Die Einflussfaktoren sind dabei erneut nach absoluten Werten absteigend geordnet.

Parameter	Wert
Net size	0,6495
Degree	0,6196
Spread x PMQG	0,3855
Net size x Degree	- 0,3449
PMQG	0,2094
Net size x PMQG	0,1201
Net size x Spread	0,0783
Degree x PMQG	0,0753
Spread	- 0,0421
Degree x Spread	- 0,0120

Tab. 24: Einfluss von Variablen und Interaktionstermen auf die TNQ<sup>835</sup>

(Quelle: Eigene Darstellung)

Der größte Unterschied zwischen den beiden Analysen entsteht bei der Betrachtung der Netzgröße, die – wie oben ausgeführt – die NQ negativ, die TNQ aber positiv beeinflusst. Der deutlich negative Wert des Interaktionsterms von Netzgröße und Gradanzahl beschreibt einen sinkenden Einfluss der Größen auf die TNQ, wenn beide gleichzeitig erhöht werden. Dies bestätigt den beschriebenen Effekt der ‚kritischen Masse‘ und erlaubt großen Netzwerken durch die Vielzahl der Verbindungen eine geringe Gradanzahl der Teilnehmer zu (über-)kompensieren.<sup>836</sup> Auffällig ist zudem die starke Abhängigkeit von Spread und PMQG. Während die PMQG als solche grundsätzlich positiv auf die TNQ wirkt, verzeichnet der Spread als einzelne Einflussgröße nahezu keinen Effekt. Beide Parameter wirken erst, wenn sie gemeinschaftlich erhöht werden. Dies ergibt sich aus der Eigenschaft des Spread, die Wahlfreiheit zu erhöhen – kleine Spreads ermöglichen, wie dargestellt, keine freie Nachbarschaftswahl – und der Eigenschaft der PMQG, die Auswahl zu verbessern. Eine Verbesserung der Auswahl ist nur in engen Grenzen möglich, wenn die Anzahl potenzieller Partner durch einen kleinen Spread stark begrenzt wird. Für die Implementierung von CroCoBIN-basierten Netzwerken ist daher grundsätzlich zumindest eine geringe ‚Freigabe‘ der Verbindungszahl zu empfehlen. Die entsprechend durch mehr

<sup>835</sup> Für die Analyse gilt  $R^2 = 0,8841$ .

<sup>836</sup> Vgl. Abschnitt 6.4.3.2.

Überführungen belasteten Teilnehmer müssen dann auf anderem Wege entlastet oder entschädigt werden.<sup>837</sup>

Der Einfluss der weiteren Variablen und Interaktionen ist gering. Es zeigt sich an der Analyse aber erneut, dass die Messgrößen (hier die NQ) nur in Abhängigkeit von bestimmenden Einflussgrößen eine Prognose der Kontrollgrößen (hier der TNQ) erlauben. Der folgende Abschnitt stellt daher abschließend dar, welche Aussagekraft MQ und NQ besitzen und wie die entsprechenden tatsächlichen Qualitäten des Netzes prognostiziert werden können.

#### 6.4.4 Aussagekraft und Prognosefähigkeit der Gütemaße

Die bisherigen Betrachtungen der Gütemaße haben gezeigt, dass ihre Verwendbarkeit nur gegeben ist, wenn sie angemessen parametrisiert werden. Für die MQ geschieht dies durch die Gewichtung der Teilqualitäten. Die PQ und RQ werden maßgeblich durch die Auswahl einer geeigneten t-Norm bestimmt und bilden dann eine NQ ab, die für unterschiedliche Netzwerkgrößen die tatsächliche Netzwerkqualität so zutreffend bestimmt, dass sich statistisch signifikante Aussagen über die Gestaltung der Netzwerke als solche treffen lassen. Die Gütemaße erfüllen damit ihren Zweck, die Qualität der Netzwerke wiederzugeben und ihre Gestaltung positiv zu beeinflussen. Die wichtigsten der beschriebenen Größen sind dabei die MQ als Qualität einer einzelnen Verbindung, da sie häufig bestimmt werden muss und in direktem Zusammenhang mit dem betriebenen Aufwand zur Peer-Paarung steht, sowie die NQ als zusammenfassende und ‚global‘ bestimmte Größe, da sie die Festlegung übergreifender Richtlinien und Netzwerkstrukturen maßgeblich beeinflusst.

Für MQ und NQ ist per Definition ein möglichst hoher Wert anzustreben, um die tatsächliche Qualität der Netzwerke auf ein hohes Niveau zu bringen. Die Zusammenhänge der einzelnen (Teil-)Qualitäten sind dabei aber nicht linear und die in der Implementierung tatsächlich verwendeten Skalen von Mess- und Kontroll-

---

<sup>837</sup> Denkbar sind entsprechende Kompensationszahlungen. Die Bestimmung dieser ist jedoch schwierig, da bei einer direkten Überführung auch immer ein zusätzlicher Nutzen für den Teilnehmer gegeben ist, der bereits viele Verbindungen aufweist, da direkte Verbindungen per Definition nicht durch indirekte Verbindungen hinsichtlich ihres Informationserhalts übertroffen werden können (vgl. Abschnitt 5.2.1). Nutzen und Aufwand müssen hier zunächst gegeneinander abgewogen werden, bevor eine Kompensation ausgehandelt werden kann. Für die Komplexität dieses Vorgehens am Beispiel von Wissensnetzwerken vgl. Weitzel/Beimborn/König (2003): 205f.



größen weichen teilweise stark voneinander ab.<sup>838</sup> Eine effektive Ressourcenallokation setzt jedoch voraus, dass der Zusammenhang der Größen zumindest näherungsweise bekannt ist, damit nicht hoher Aufwand betrieben wird, um eine Messgröße zu erhöhen, die durch eine große Wertänderung auch eine hohe Änderung der tatsächlichen Qualität suggeriert, obwohl der Grenznutzen weiterer Erhöhungen durch den tatsächlichen funktionalen Zusammenhang von Messgröße und Kontrollgröße ggf. nur gering ist.<sup>839</sup>

Für die TMQ lässt sich unter Berücksichtigung des dargestellten Zusammenhangs<sup>840</sup> auf Basis der MQ eine Schätzung ermitteln. Dabei wird erneut auf die Methode der linearen Regression zurückgegriffen. Um den nicht-linearen Zusammenhang abzubilden, wird neben der MQ auch die quadrierte MQ als Funktionsterm abgebildet, d. h. die TMQ als abhängige Variable der Form  $TMQ = \beta_1 \times MQ^2 + \beta_2 \times MQ + u$  mit  $u$  als linearer Verschiebung der Regressionskurve betrachtet.<sup>841</sup> Es ergibt sich damit als geschätzte tatsächliche Mapping-Qualität:

$$\widehat{TMQ} = 4,5270 \times MQ^2 - 4,8000 \times MQ + 1,2561$$

Das Bestimmtheitsmaß der angegebenen Regression liegt bei  $R^2 = 0,9861$  und erlaubt damit eine sehr genaue Prognose der tatsächlich erzielten TMQ, die noch verbessert werden kann ( $R^2 = 0,9934$ ), wenn der Regressionsterm ein Polynom dritter Ordnung darstellt:

$$\widehat{TMQ} = 5,7161 \times MQ^3 - 7,600 \times MQ^2 + 3,3930 \times MQ + 0,5026$$

Die Veränderung des Bestimmtheitsmaßes ist dabei nur gering, jedoch erlaubt die dritte Ordnung die Vermeidung der für Polynome zweiter Ordnung typische Para-

<sup>838</sup> Vgl. die Hinweise zur Nicht-Linearität von MQ und TMQ in Abschnitt 6.4.1 sowie die Darstellung der stark abweichenden Wertbereiche von NQ und TNQ in Abschnitt 6.4.3.2.

<sup>839</sup> Der *Grenznutzen* beschreibt allgemein den Nutzen, den die letzte eingesetzte Einheit eines Gutes erbringt (vgl. Söllner (2015): 46). Deutlich wird die Problematik hier am Beispiel der MQ, deren Einfluss auf die TMQ sehr gering ist (vgl. Abschnitt 6.4.1) und die deshalb sehr gering gewichtet wird. Eine Erhöhung der Konsistenz bei einem Mapping führt zu einer deutlich geringeren Erhöhung der TMQ.

<sup>840</sup> Vgl. Abschnitt 6.4.1, insbesondere Abb. 58.

<sup>841</sup> Vgl. zum allgemeinen Vorgehen mit quadratischen Termen die Hinweise zu Interaktionstermen in Abschnitt 6.4.3.3 sowie Wooldridge (2013): 186f.

bel, die für sehr geringe MQ-Werte fälschlicherweise hohe und für MQ-Werte um 0,55 negative TMQ-Werte ausweist (vgl. Abb. 76).<sup>842</sup>

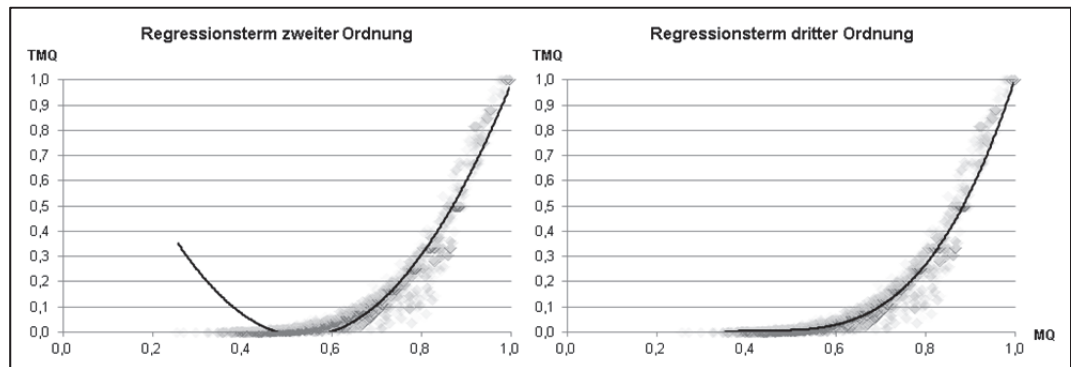


Abb. 76: TMQ und geschätzte TMQ in Abhängigkeit von der MQ  
(Quelle: Eigene Darstellung)

Die Schätzung der TNQ auf Basis der NQ erfordert – wie in Abschnitt 6.4.3.3 herausgearbeitet – zumindest die Berücksichtigung der Netzgröße, da diese maßgeblich die Achsenabschnitte von NQ und TNQ in ihren jeweiligen Wertebereichen beeinflusst.<sup>843</sup> Abb. 77 stellt die durchschnittlich erzielten TNQ-Werte unterschiedlicher NQ-Werte und Netzwerkgrößen dar. Es lässt sich auf dieser Basis eine erste Schätzung der für ein gegebenes Netzwerk erzielten TNQ ermitteln, sofern NQ und Netzwerkgröße im simulierten Parameterbereich liegen.

<sup>842</sup> Die so durchgeführte Schätzung ist durch die zuvor durchgeführte Auswahl geeigneter Werte für den Gewichtungsvektor bereits in der zuerst vorgestellten Variante präziser als eine Regression auf den quadratischen Termen der Teilqualitäten ( $R^2 = 0,9773$ ).

<sup>843</sup> Vgl. die Darstellung von TNQ und NQ für unterschiedliche Netzwerktypen und -größen in Abb. 70.

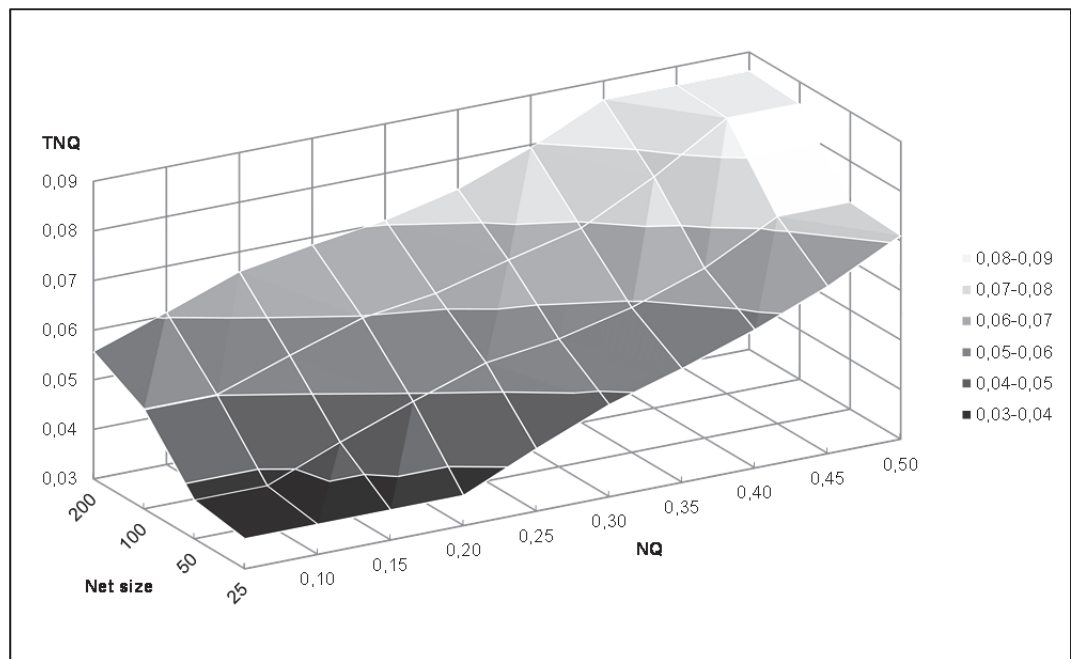


Abb. 77: 3D-Plot der TNQ von Zufallsnetzwerken in Abhängigkeit von Netzgröße und NQ

(Quelle: Eigene Darstellung)

Zur Durchführung einer präziseren Schätzung lässt sich auch für NQ und TNQ mit Hilfe der Regressionsanalyse ein funktionaler Zusammenhang ermitteln. Es zeigt sich bei der Betrachtung der TNQ in Abhängigkeit von NQ und Netzwerkgröße jedoch eine deutliche, durch den hohen Einfluss der Gradanzahl verursachte ‚Streifen‘- oder ‚Gruppen‘-Bildung innerhalb der einzelnen Netzwerkgrößen.<sup>844</sup> Abb. 78 stellt diese durch die aufgeteilte Ansicht der TNQ- und NQ-Werte für einzelne Netzwerkgrößen deutlich heraus. Die Ergebnisse einer Netzwerkgrößen- und Gradanzahl-abhängigen Schätzung ergeben daher erheblich präzisere Werte als eine Schätzung, die ausschließlich die Netzgröße als Parameter neben der NQ berücksichtigt. Tab. 25 stellt die Bestimmtheitsmaße der Gradanzahl-abhängigen und Gradanzahl-unabhängigen Regressionsanalyse einander gegenüber.

<sup>844</sup> Vgl. dazu auch die Hinweise und Berechnungen in den Abschnitten 6.4.3.2 und 6.4.3.3 sowie insbesondere die Darstellung der TNQ in Abhängigkeit von der NQ für Zufallsnetzwerke in Abb. 70 (oben links).

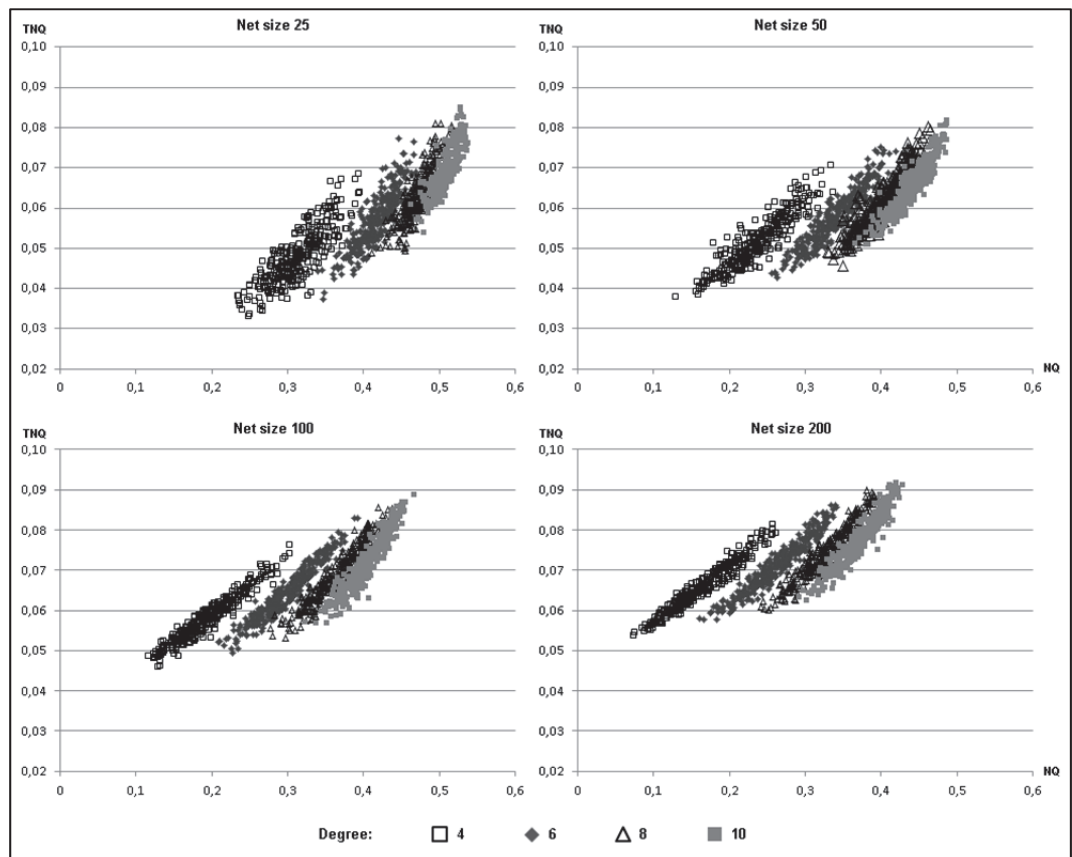


Abb. 78: TNQ von Zufallsnetzwerken in Abhängigkeit von NQ und Gradanzahl für unterschiedliche Netzgrößen  
(Quelle: Eigene Darstellung)

Größe	Gradanzahl					
	Unabhängig	Ø Abhängig	4	6	8	10
25	0,7893	0,7042	0,6972	0,7324	0,7109	0,6763
50	0,6954	0,8538	0,8315	0,8649	0,8671	0,8517
100	0,7302	0,9126	0,9256	0,9249	0,9142	0,8855
200	0,7003	0,9232	0,9467	0,9294	0,9246	0,8921

Tab. 25: Bestimmtheitsmaße Gradanzahl-abhängiger und -unabhängiger Regressionsanalysen von TNQ und NQ  
(Quelle: Eigene Darstellung)

Eindeutig erkennbar ist das für Netzgrößen  $n \geq 50$  deutlich höhere Bestimmtheitsmaß der Gradanzahl-abhängigen Regressionsgeraden. Für  $n = 25$  zeigt sich eine leichte Verschlechterung des Wertes, die auf die vergleichsweise starke Streuung der NQ-Werte – insbesondere für kleine Gradanzahlen – zurückzuführen ist, was durch die Reduktion der betrachteten Datenpunktzahl bei Auf-

teilung dieser nach Gradanzahlen zu einer höheren Varianz in Bezug auf die ermittelte Regressionsgerade führt.<sup>845</sup> Zur Ermittlung einer möglichst allgemeingültigen Schätzfunktion wird entsprechend die Gradanzahl-abhängige Methode verwendet. Dazu wurden mehrere Regressionsanalysen kombiniert, um Netzwerkgrößen- und Gradanzahl-spezifisch Steigung und Achsenabschnitt der sechzehn einzeln zu betrachtenden Datenpunktmengen (vier Größen zu je vier Gradanzahlen) als Einflussparameter einer gesamtheitlich anwendbaren Funktion zu ermitteln. Insbesondere der Verdopplung der betrachteten Netzwerkgrößen wurde dabei durch Verwendung logarithmischer Terme Rechnung getragen.<sup>846</sup> Es ergibt sich damit für die geschätzte TNQ (mit  $n$  als Größe und  $d$  als Gradanzahl des Netzwerks):

$$\widehat{TNQ} = \left( 0,0241 \times \log_2 \frac{n}{25} + 0,0786 \times \log_2 \frac{d}{4} + 0,1866 \right) \times NQ \\ + \left( 0,0239 \times \log_2 \frac{n}{25} - 0,0467 \times \log_2 \frac{d}{4} - 0,0140 \right)$$

Sowohl für die geschätzte TMQ als auch für die geschätzte TNQ ist dabei zu berücksichtigen, dass die ermittelten Werte auf Basis der simulierten Werte und der dabei getroffenen Annahmen über das Modell ermittelt sind. Sie besitzen daher – insbesondere in Bezug auf die ermittelten Koeffizienten – nur für das vorgestellte Modell oder ihm ähnliche Implementierungen Gültigkeit, vermitteln aber einen Eindruck von den grundsätzlich erwartbaren funktionalen Zusammenhängen.

Es ist zudem zu erwarten, dass sich die  $\widehat{TMQ}$  gegenüber Veränderungen im Modell robust verhält, solange die Gewichtung der Teilqualitäten erhalten bleibt, da der Zusammenhang von Mess- und Kontrollgröße auch bei Veränderung des Matching-Verfahrens in Bezug auf den wichtigsten Einflussfaktor – die  $MQ_s$  – durch das ‚einfache‘ Abzählen der direkt übertragbaren Elemente weitgehend bestehen bleibt. Die  $\widehat{TNQ}$  hingegen ist stark abhängig von dem Wertebereich der NQ, der insbesondere bei hohen durchschnittlichen Pfadlängen maßgeblich von der einge-

<sup>845</sup> Es zeigt sich hier erneut das bereits herausgearbeitete abweichende Verhalten kleiner Netzwerke in Bezug auf die überprüften Qualitäten (vgl. Abschnitte 6.4.3.2 und 6.4.3.3).

<sup>846</sup> Das Vorgehen ist in Anhang 5 ausführlich dargestellt. Die Verwendung logarithmischer Terme ist im Rahmen der multiplen linearen Regression üblich zur Behandlung nicht-linearer Abhängigkeiten (vgl. Wooldridge (2013): 183). Durch ihre Anwendung steigt hier das Bestimmtheitsmaß der Regression deutlich von  $R^2 = 0,8335$  auf  $R^2 = 0,9219$ . Nur zum Vergleich sei hier das Bestimmtheitsmaß der logarithmierten, Gradanzahl-unabhängigen Regression angegeben. Es liegt bei  $R^2 = 0,8963$ , was die Berücksichtigung der Gradanzahl auch statistisch stützt.

setzten t-Norm beeinflusst wird.<sup>847</sup> Zudem kann eine Übertragbarkeit der Schätzung auf deutlich abweichende Parameter (Netzgröße, Gradanzahl) nicht gewährleistet werden.<sup>848</sup> Das hohe Bestimmtheitsmaß der Schätzungen bestätigt jedoch durch die positive Steigung der Regressionsgeraden den positiven Zusammenhang von NQ und TNQ. Sowohl für die in der Schätzfunktion verwendeten Parameter (Größe, Gradanzahl) als auch für die nicht verwendeten Parameter (PMQG, Spread) lässt sich daher ein eindeutiger, positiver Zusammenhang zur TNQ bestimmen, sodass diese gemäß der in Abschnitt 6.4.3.3 ermittelten Größe ihres Einflusses in implementierten CroCoBIN-basierten Netzwerken verbessert werden sollten, um die tatsächliche Netzqualität zu heben. MQ und NQ erfüllen damit den ihnen zugedachten Zweck, für Anwender des Netzwerks ermittelbare Größen darzustellen, die Aufschluss über die tatsächlich vorhandenen Qualitäten des Netzwerks geben.

#### 6.4.5 Abschließende Betrachtungen der Ergebnisse und des Fallbeispiels

Die durchgeführten Simulationen haben gezeigt, dass sich der Umfang der Parameter und ihrer Wertebereiche merklich einschränken lässt, um sinnvolle Netzwerk-Konfigurationen zu erhalten. Tab. 26 stellt in Analogie zu der einführenden Übersicht in Abschnitt 6.4 abschließend die relevanten Ergebnisse der einzelnen, nach Abschnitten geordneten Simulationen dar.

Abschnitt	Betrachtete Größe(n)	Ergebnis
6.4.1	MQ, MQ <sub>s</sub> , MQ <sub>c</sub> , MQ <sub>b</sub>	Die MQ wird durch einen Gewichtungsvektor $(MQ_s; MQ_c; MQ_b) = (0,60; 0,05; 0,35)$ sinnvoll abgebildet.
6.4.2	RQ (NQ), TRQ (TNQ)	Die t-Norm $T_{1,25}^Y$ führt zu sehr guten Ergebnissen der Messgrößen und zudem zu einer starken Korrelation dieser mit den Kontrollgrößen.
6.4.3.1	NQ	Die PMQG führt nur im Bereich $[0,4; 0,8]$ zu merklichen Veränderungen, der Parameter $\mu$ hat nahezu keine Auswirkungen. Der ‚Selfish‘-Parameter führt in der Konfiguration <i>false</i> zu starken Veränderungen der NQ.

<sup>847</sup> Vgl. die stark unterschiedlichen Abdeckungen der Wertebereiche verschiedener t-Normen in Abschnitt 5.3.2.

<sup>848</sup> Vgl. hierzu auch die Ausführungen zu sehr großen Netzwerken in Abschnitt 7.2 und Anhang 6.

Abschnitt	Betrachtete Größe(n)	Ergebnis
6.4.3.2	NQ, TNQ	Netzwerke mit ‚selfish‘-Konfiguration erzielen für beide Netzwerktypen im Durchschnitt höhere Werte der TNQ als solche mit ‚non-selfish‘-Konfiguration.  Zufallsnetzwerke mit ‚selfish‘-Konfiguration erzielen zudem höhere Werte der TNQ als skalenfreie Netzwerke gleicher Konfiguration.  Die TNQ steigt mit der Netzwerkgröße für Netzwerke mit $n \geq 50$ .
6.4.3.3	NQ, TNQ	Den größten Einfluss auf NQ und TNQ weisen die Netzwerkgröße und die Gradanzahl auf.  Insbesondere die TNQ wird auch von Spread und PMQG beeinflusst, hervorzuheben ist hier vor allem der entsprechende Interaktionsterm ‚Spread x PMQG‘, der positiv mit der TNQ korreliert.
6.4.4	MQ, TMQ, NQ, TNQ (Schätzer)	Für die Schätzer von TMQ und TNQ lassen sich sehr gute ( $R^2 > 0,92$ ) Funktionen bestimmen, die sich als Polynom dritter Ordnung ( $\widehat{TMQ}$ ) bzw. kombinierte, teillogarithmierte Regressionsgerade ( $\widehat{TNQ}$ ) darstellen lassen.

Tab. 26: Zusammenfassung der Ergebnisse der einzelnen Simulationen

(Quelle: Eigene Darstellung)

Vor dem Hintergrund des insbesondere in den Abschnitten 2.4, 3.3.4 und 6.2 diskutierten Fallbeispiels präsentiert sich CroCoBIN mit der einfachen MQ-Errechnung, der als nützlich und nutzbar identifizierten t-Norm und der Eignung der NQ als verwendbarer Netzwerk-Messgröße als geeigneter Umsetzungsrahmen für unternehmensübergreifende BI-Netzwerke. Als positiv für den konkreten Anwendungsfall bzw. das Fallbeispiel kann dabei die Erkenntnis gelten, dass größere Netzwerke nicht zu sinkenden, sondern – im Gegenteil – steigenden tatsächlichen Netzqualitäten führen. Die Aufnahme einer größeren Anzahl an Kooperationspartnern ist damit für kleinere Unternehmen weniger problematisch als vielmehr gewinnversprechend. Auch zeigt sich, dass eine zentrale Stelle oder ein zentrales Schema nicht notwendig sind. Gemessen an den Qualitäten vollständig verbundener Netze (deren 1:1-Überführung je Teilnehmer immer mindestens so gut ist wie die Über- und Rückführung von Daten in ein zentrales Schema) führen bereits Netzwerke mit geringen Gradzahlen zu sehr guten tatsächlichen Netzwerkqualitäten.<sup>849</sup>

<sup>849</sup> Vgl. Abschnitte 6.4.3.2 und 6.4.4.

## **7 Resümee und Ausblick**

Die vorliegende Arbeit stellt eine ausführliche Betrachtung unternehmensübergreifender BI-Netzwerke und deren Güte(-Messung) dar. Die folgenden Abschnitte rekapitulieren Gang und Erkenntnisse der Arbeit und zeigen das Potenzial weiterer Forschung in diesem Bereich auf.

### **7.1 Zusammenfassung der Arbeit**

Basierend auf einem gestaltungsorientierten Forschungsansatz wurden zunächst in Kapitel 1 Motivation und Ziele der Arbeit formuliert, die verwendeten Forschungsmethoden erläutert und diese in den Kontext der üblichen Vorgehensweisen in der deutschen und internationalen Wirtschaftsinformatik gestellt.

In Kapitel 2 wurde ein Überblick verschiedener Ansätze zur Unternehmenskooperation und darauf anwendbarer Klassifikationskriterien gegeben. Als wichtige Konzepte ließen sich horizontal ausgerichtete Strategische Allianzen und das vertikal ausgerichtete Supply Chain Management identifizieren. Basierend auf den Erkenntnissen eines real existierenden Netzwerks kleiner und mittlerer Unternehmen der Automobilindustrie wurde ein Fallbeispiel entwickelt, das gemeinschaftliche Entscheidungsunterstützung im Bereich F&E beschreibt.

Die Konzepte zur Entscheidungsunterstützung wurden im folgenden Kapitel 3 erläutert und in ihren (historischen) Zusammenhang gesetzt. Business Intelligence, Data-Warehouse-Konzepte und Gruppenentscheidungsunterstützung wurden als Hauptthemen erörtert, die aktuelle Diskussion der als ‚Collaborative Business Intelligence‘ bezeichneten Entwicklungen einer ausführlichen Literaturrecherche unterzogen. Es ließen sich drei Hauptrichtungen der CBI herausstellen, die interne Kommunikation, partnerschaftliche Datenhaltung und partnerschaftliche Analyse fokussieren.

Zur Abgrenzung von unternehmensinternen Ansätze wurde in Kapitel 4 ‚Cross-Company Business Intelligence‘ als Begriff und Schwerpunkt dieser Arbeit eingeführt und definiert. Im Folgenden wurden systematisch Ansätze zur Abbildung paralleler und verteilter Datenverarbeitung sowie der Harmonisierung und Überführung unterschiedlicher Datenbestände vorgestellt und hinsichtlich ihrer Eignung für den BI-Kontext analysiert. Einen wesentlichen Bestandteil nahm dabei die Betrachtung und Bewertung geeigneter Netzwerkformen, insbesondere der



P2P-Netzwerke, ein. Als Zwischenergebnis der Arbeit wurde mit ‚CroCoBIN‘ ein Referenzmodell zur Abbildung eines unternehmensübergreifenden P2P-basierten BI-Netzwerks entwickelt, das fachliche, technische und organisatorische Komponenten vereint und Basisanforderungen für die Anwendung definiert. Anhand des Modells wurden Forschungslücken insbesondere in den Punkten ‚Nachbarschaftswahl‘ und ‚Netzwerkparametergestaltung‘ sowie der zwingend notwendigen Gütemessung identifiziert.

Auf dieser Basis wurde in Kapitel 5 zunächst der Qualitätsbegriff im Informations-Kontext beschrieben und genutzt, um unter Berücksichtigung bestehender Ansätze und Kriterien Mess- und Einflussgrößen in CroCoBIN-basierten Netzwerken herauszuarbeiten. Für diese wurden nachfolgend anwendbare Mess- und Kontrollgrößen entwickelt. Im Fokus standen dabei gleichermaßen die praktische Nutzbarkeit der Größen sowie deren Fähigkeit, die tatsächlichen Qualitäten exakt zu beschreiben und durch eine klare Parameterdefinition deren Erhöhung zu ermöglichen.

Um die Anwendbarkeit der ermittelten Größen nachzuweisen, wurde in Kapitel 6 die für die Arbeit erstellte prototypische Implementierung eines CroCoBIN vorgestellt und evaluiert. Diese stellt für alle identifizierten Einflussgrößen die Möglichkeit zur Parametrisierung von Simulationen bereit. Sie lieferte umfangreiche Datensätze zur Analyse der komplexen Zusammenhänge in den betrachteten Netzwerken. Die Validität der Simulationen wurde erfolgreich geprüft, sodass die Datensätze als Grundlage ausführlicher Analysen unterschiedlicher Netzwerkkonfigurationen verwendet werden konnten. Diese ergaben, dass Netzwerke im betrachteten Umfang nicht nur, wie erwartbar, von einer hohen Verknüpfungszahl und guten Prognosequalität profitieren, sondern dass auch ein Effekt der ‚kritischen Masse‘ eintritt, der für große Netzwerke eine Erhöhung der tatsächlichen Qualität beschreibt. Abschließend wurden Prognoseinstrumente entwickelt, die es erlauben, auf Basis der messbaren Größen die tatsächliche Netzwerkqualität zu bestimmen und so eine effektive Ressourcenallokation in der Netzwerkgestaltung unterstützen.

Das Erkenntnisziel der Arbeit lag in der Bestimmung der grundsätzlichen Eignung von P2P-basierten Netzwerken für unternehmensübergreifende BI-Lösungen sowie – bei Vorliegen dieser – in der Beschreibung von Einflussgrößen und Konfigurationen, die eine praktische Anwendbarkeit der Netzwerke ermöglichen. Es

wurde durch die Erkenntnisse am Referenzmodell sowie die aus den Gütemaßen abgeleiteten Gestaltungsempfehlungen erfüllt.

Als notwendige Voraussetzung dafür dienten die zur Erreichung des Gestaltungsziels der Arbeit von einer umfangreichen Literaturanalyse gestützte Entwicklung des Referenzmodells sowie die Erstellung von CroCoSIM als prototypischer Implementierung, die relevante Modellbereiche abbildet und eine parametrisierbare Simulation unternehmensübergreifender BI-Netzwerke ermöglicht. Auch das Gestaltungsziel der Arbeit wurde damit erreicht.

## **7.2 Schlussfolgerungen aus den Untersuchungen und Implikationen auf die weitere Forschung**

Die vorliegende Arbeit liefert einen umfassenden Überblick des Stands der Forschung zu unternehmensübergreifender Business Intelligence, den grundlegenden Konzepten sowie umliegenden Fachgebieten, deren Ansätze einen Beitrag zur weiteren Entwicklung des Themas bieten können.

Das im Rahmen dieser Arbeit entwickelte Referenzmodell greift diese auf, stellt aber eine nur abstrakte Sichtweise auf CCBI dar. Zur genaueren Betrachtung einzelner Aspekte kann CroCoBIN in zukünftigen Arbeiten detaillierter ausgearbeitet und ggf. in Bezug zu Branchen- oder Unternehmensspezifika gesetzt werden.<sup>850</sup>

Die vorliegende Arbeit fokussiert zudem auf einzelne Aspekte des Modells, die für die identifizierten Forschungslücken von besonderer Relevanz sind. Insbesondere hinsichtlich der organisatorischen Rahmenbedingungen besteht jedoch für die konkrete Implementierung von CroCoBIN in Unternehmen Forschungsbedarf, um eine Anwendbarkeit im Praxisfall auch längerfristig zu sichern. Dazu zählen neben Sicherheits- und Vertrauensaspekten auch die Fragestellung nach Verbindungsvielfalt und Kompensationen für besonders aktive Netzteilnehmer.<sup>851</sup>

Die weiteren Ausführungen zur Gestaltung der Netzwerke konzentrieren sich in der vorliegenden Arbeit auf einzelne, besonders schwach durchdrungene Forschungsfelder, insbesondere die Entwicklung von Gütemaßen. Bereits unter An-

---

<sup>850</sup> Vgl. hierzu auch die allgemeinen Hinweise zur Entwicklung und Anwendung von Referenzmodellen in Abschnitt 4.5.1.

<sup>851</sup> Vgl. Abschnitt 6.4.4.

wendung des bestehenden CroCoBIN-Referenzmodells zeigt die Literaturrecherche aber weitere, noch nicht ausreichend diskutierte Fragestellungen auf.<sup>852</sup>

Zu diesen zählt unter anderem die Verbesserung der Matching-Verfahren heterogener (multidimensional strukturierter) Datenbestände. Bisherige Verfahren liefern noch keine zufriedenstellenden Ergebnisse in Bezug auf eine mögliche Vollautomatisierung der Überführungen, was einen hohen Aufwand zur Peer-Paarung bedeutet und sowohl die Anzahl der Verbindungen als auch die Schnelligkeit der Wechsel von Partnern im Netzwerk negativ beeinflusst. Neben einer Verbesserung der Verfahren selbst sind in diesem Zusammenhang auch deren Auswirkungen zu diskutieren. Die vorgestellten Analysen erlauben zwar die für P2P-Netzwerke charakteristische ‚lockere‘ und unabhängige Bindung der Partner, es wurden jedoch keine Tests der Netzwerkdynamik, d. h. des Verhaltens bei sehr hoher Fluktuation der Teilnehmer durchgeführt. Zukünftige Forschungsarbeiten sollten diesen Aspekt genauso berücksichtigen wie die Fragestellung, ob ggf. der Einsatz von automatischen Matching-Verfahren die Anzahl der Verbindungen in einem so großen Maße erhöhen kann, dass die gesteigerte durchschnittliche Gradanzahl in den Netzen den negativen Effekt nicht optimaler (i. S. v. maximal durch menschlich durchgeführtes Matching erzielbarer) Übereinstimmung kompensiert.

Nur als Randbedingung für ‚gute‘ Netzwerke wurde in dieser Arbeit die Robustheit der Netzwerke geprüft. Die durchgeführten Simulationen weisen darauf hin, dass die im Allgemeinen als robuster geltenden skalenfreien Netzwerke für den konkreten Anwendungsfall keine erkennbaren Vorteile generieren und zudem schlechtere Netzwerkqualitäten aufweisen. Hier können zukünftige Arbeiten die Gründe für dieses Verhalten evaluieren und ggf. in Verbindung mit anderen Topologien grundlegende Verbesserungen der Netzwerkstruktur entwickeln.

Variabilität und Robustheit der Netzwerke sollten umso mehr betrachtet werden, als die Netzwerkgröße einen entscheidenden Einfluss auf alle Netzwerkqualitäten sowie deren Ermittlung und Prognose aufweist. Der Effekt der ‚kritischen Masse‘ konnte für mittlere bis große Netzwerke gezeigt werden. Eine ergänzende Studie weist jedoch darauf hin, dass der Effekt nicht für beliebig große Teilnehmerzahlen anhält.<sup>853</sup> Während etwa 300 bis 350 Teilnehmer noch eine höhere TNQ verursa-

---

<sup>852</sup> Vgl. die Darstellung von CroCoBIN-Bestandteilen und Inhalten der untersuchten Publikationen in Abschnitt 4.5.3, insbesondere Tab. 10.

<sup>853</sup> Vgl. Anhang 6.

chen, sinkt diese bei weiterer Vergrößerung des Netzes wieder auf das Niveau von etwa 100 Teilnehmern ab, beginnt bei mehr als 750 Teilnehmern jedoch wieder zu steigen.<sup>854</sup> Die entsprechenden Zahlen sind jedoch modellgebunden und müssen für den tatsächlich implementierten Fall ggf. neu ermittelt werden.

Abschließend ist das vorgestellte Referenzmodell an realitätsnahen, aber nicht realen Daten evaluiert worden. Eine Implementierung der Technologie in Unternehmen ist daher – auch zur Untersuchung der tatsächlich eintretenden ökonomischen Effekte – eine erstrebenswerte Leistung der weiteren wissenschaftlichen und praktischen Ausarbeitung des Forschungsfeldes. Hier kann auch die Verwendung stark heterogener Datenstrukturen, hochdimensionaler Datenbestände und komplexer Überführungsverfahren weitere Erkenntnisse generieren. Die vorliegende Arbeit bietet dabei ein wohldefiniertes und erprobtes Rahmenwerk als Grundlage der Untersuchung dieser und einer Vielzahl weiterer betriebswirtschaftlich relevanter Fragestellungen in den Bereichen Kooperation und Entscheidungsunterstützung.

---

<sup>854</sup> Das veränderte Verhalten verhindert auch die durchgängige Anwendung der ermittelten Prognosemaße, deren Bestimmtheitsmaß für  $n = 300$  auf  $R^2 = 0,7721$  und für  $n = 500$  bereits auf  $R^2 = 0,5012$  sinkt.

## Literaturverzeichnis

- Abelló, Alberto/Romero, Oscar/Pedersen, Torben/Berlanga Llavori, Rafael/Nebot, Victoria/Aramburu, María J./Simitsis, Alkis* (2014): Using Semantic Web Technologies for Exploratory OLAP: A Survey. In: IEEE Transactions on Knowledge and Data Engineering (Early Access): 1–20.
- Abels, Sven/Haak, Liane/Hahn, Axel* (2005): Identification of common methods used for ontology integration tasks. In: Hahn, Axel (Hrsg.): Proceedings of the first international workshop on Interoperability of heterogeneous information systems (IHIS '05). New York, NY: ACM: 75–78.
- Aberer, Karl/Alima, Luc O./Ghodsi, Ali/Girdzijauskas, Saruns/Haridi, Seif/Hauswirth, Manfred* (2005): The essence of P2PA a reference architecture for overlay networks. In: Caronni, Germano (Hrsg.): Proceedings of the Fifth IEEE International Conference on Peer-to-Peer Computing. Los Alamitos, CA: IEEE: 11–20.
- Agrawal, Rakesh/Srikant, Ramakrishnan/Thomas, Dilys* (2005): Privacy preserving OLAP. In: Ozcan, Fatma (Hrsg.): Proceedings of the 2005 ACM SIGMOD international conference on Management of data (SIGMOD '05). New York, NY: ACM: 251–262.
- Ahlf, Henning* (2013): Identifikation von Influentials in virtuellen sozialen Netzwerken: Eine agentenbasierte Modellierung und Simulation sozialer Beeinflussungsprozesse. Duisburg: Universität Duisburg-Essen.
- Akinde, Michael O./Böhlen, Michael H./Johnson, Theodore/Lakshmanan, Laks V. S./Srivastava, Divesh* (2003): Efficient OLAP query processing in distributed data warehouses. In: Information Systems, 28 (1-2): 111–135.
- Alanazi, Jazem/Chatfield, Akemi* (2013): Sharing Government-Owned Data with the Public: A Cross-Country Analysis of Open Data Practice in the Middle East. Paper 16. In: Association for Information Systems (Hrsg.): Proceedings of the 18th Americas Conference on Information Systems 2012 (AMCIS). Red Hook, NY: Curran.
- Albert, Reka/Jeong, Hawoong/Barabasi, Albert-Laszlo* (2000): Error and attack tolerance of complex networks. In: Nature, 406 (6794): 378–382.
- Albert, Réka/Barabási, Albert-László* (2002): Statistical mechanics of complex networks. In: Reviews of Modern Physics, 74 (1): 47–97.
- Albrecht, Jens/Lehner, Wolfgang* (1998): On-line Analytical Processing in Distributed Data Warehouses. In: IEEE (Hrsg.): Proceedings of the International Database Engineering and Applications Symposium (IDEAS'98). Los Alamitos, CA: IEEE: 78–85.
- Alicke, Knut* (2005): Unternehmensübergreifendes Supply Chain Management. In: Alicke, Knut (Hrsg.): Planung und Betrieb von Logistiknetzwerken. 2. Aufl. Berlin, Heidelberg: Springer: 1–208.
- Aligon, Julien/Gallinucci, Enrico/Golfarelli, Matteo/Marcel, Patrick/Rizzi, Stefano* (2014a): A Collaborative Filtering Approach for Recommending OLAP Sessions. In: Decision Support Systems, In Press.

- Aligon, Julien/Golfarelli, Matteo/Marcel, Patrick/Rizzi, Stefano/Turricchia, Elisa* (2014b): Similarity measures for OLAP sessions. In: *Knowledge and Information Systems*, 39 (2): 463–489.
- Alter, Steven* (1980): *Decision support systems*. Reading, MA et al.: Addison-Wesley.
- Alwis, Dee* (2004): The role of intellectual capital in organisational value creation: An application of a theoretical model to two case studies. In: Brunel University London (Hrsg.): *Dept of Computer Science Theses*. London: Brunel University London. URL: <http://bura.brunel.ac.uk/handle/2438/5068>, Abruf am 28.11.2014.
- Angelov, S./Grefen, P./Greefhorst, D.* (2009): A classification of software reference architectures: Analyzing their success and effectiveness. In: IEEE (Hrsg.): *Proceedings of the European Conference on Software Architecture (WICSA/ECSA 2009)*: 141–150.
- Aouiche, Kamel/Jouve, Pierre-Emmanuel/Darmont, Jérôme* (2006): Clustering-Based Materialized View Selection in Data Warehouses. In: Manolopoulos, Yannis/Pokorný, Jaroslav/Sellis, Timos (Hrsg.): *Advances in Databases and Information Systems*. Berlin, Heidelberg, New York: Springer: 81–95.
- Arshinder, Kaur/Kanda, Arun/Deshmukh, S. G.* (2011): A review on supply chain coordination. In: Choi, Tsan-Ming/Cheng, T.C. Edwin (Hrsg.): *Supply chain coordination under uncertainty*. Berlin, Heidelberg: Springer: 39–82.
- Association for Information Systems* (2011): Senior Scholars' Basket of Journals. URL: <http://start.aisnet.org/?SeniorScholarBasket>, Abruf am 11.05.2015.
- Audimoolam, Srinivasaragavan/Dutta, Rick* (2005): Decision support system for supply chain management. Patentnr. US 2005/0209732 A1.
- Auer, Benjamin R./Seitz, Franz* (2013): *Grundkurs Wirtschaftsmathematik*, 4. Aufl. Wiesbaden: Gabler.
- Averbuch, Jakob* (2013): Enhanced usage of business intelligence systems. Patentnr. US 2013/0018840.
- Baars, Henning/Felden, Carsten/Gluchowski, Peter/Hilbert, Andreas/Kemper, Hans-Georg/Olbrich, Sebastian* (2014): Gestaltung der nächsten Inkarnation von Business Intelligence. In: *WIRTSCHAFTSINFORMATIK*, 56 (1): 13–19.
- Backhaus, Klaus* (2011): *Multivariate Analysemethoden*, 13. Aufl. Berlin, Heidelberg: Springer.
- Baker, Michael J.* (2000): Writing a literature review. In: *The Marketing Review*, 1 (2): 219–247.
- Balling, Richard* (1998): *Kooperation*, 2. Aufl. Frankfurt am Main et al.: Lang.
- Bamberg, Günter/Coenenberg, Adolf G./Krapp, Michael* (2008): *Betriebswirtschaftliche Entscheidungslehre*, 14. Aufl. München: Vahlen.
- Banek, Marko/Vrdoljak, Boris/Tjoa, A. Min/Skocir, Zoran* (2007): Automating the Schema Matching Process for Heterogeneous Data Warehouses. In: Song, Il Y./Eder, Johann/Nguyen, Tho M. (Hrsg.): *Proceedings of the 9th International Conference on Data Warehousing and Knowledge Discovery (DAWAK)*. Berlin, Heidelberg: Springer: 45–54.

- Bankhofer, Udo/Vogel, Jürgen* (2008): Datenanalyse und Statistik. Wiesbaden: Gabler.
- Banks, Jerry* (1998): Handbook of simulation. New York et al.: Wiley.
- Barabasi, Albert-László./Albert, Réka* (1999): Emergence of scaling in random networks. In: Science, 286 (5439): 509–512.
- Barabási, Albert-László/Bonabeau, Eric* (2003): Scale-Free Networks. In: Scientific American, 288 (5): 50–59.
- Barkai, David* (2001): Peer-to-peer computing. Hillsboro, OR: Intel Press.
- Bass, Len/Clements, Paul/Kazman, Rick* (2009): Software architecture in practice, 2. Aufl. Boston et al.: Addison-Wesley.
- Batista, Maria da Conceição Moraes/Salgado, Ana C.* (2007): Information Quality Measurement in Data Integration Schemas. In: Koch, Christoph/Gehrke, Johannes/Garofalakis, Minos N./Srivastava, Divesh/Aberer, Karl/Deshpande, Anand/Florescu, Daniela/Chan, Chee Y./Ganti, Venkatesh/Kanne, Carl-Christian (Hrsg.): Proceedings of the 33rd international conference on very large data bases (VLDB 2007). Wien: ACM: 61–72.
- Bauer, Andreas/Günzel, Holger* (2013): Data-Warehouse-Systeme, 4. Aufl. Heidelberg: Dpunkt.
- Beck, Dieter* (2001): Sozialpsychologie kollektiver Entscheidungen. Wiesbaden: Westdeutscher.
- Becker, Jörg/Holten, Robert/Knackstedt, Ralf/Niehaves, Björn* (2004): Epistemologische Positionierungen in der Wirtschaftsinformatik am Beispiel einer konsensorientierten Informationsmodellierung. In: Frank, Ulrich (Hrsg.): Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Wiesbaden: DUV: 335–366.
- Becker, Jörg/Holten, Roland/Knackstedt, Ralf/Niehaves, Björn* (2003): Forschungsmethodische Positionierung in der Wirtschaftsinformatik – epistemologische, ontologische und linguistische Leitfragen. Arbeitsbericht des Instituts für Wirtschaftsinformatik Nr. 93. Westfälische Wilhelms-Universität Münster, Institut für Wirtschaftsinformatik, Münster.
- Becker, Jörg/Knackstedt, Ralf* (2003): Konstruktion und Anwendung fachkonzeptioneller Referenzmodelle im Data Warehousing. In: Uhr, Wolfgang/Esswein, Werner/Schoop, Eric (Hrsg.): Wirtschaftsinformatik 2003: Medien - Märkte - Mobilität. Heidelberg: Physica: 415–434.
- Becker, Jörg/Niehaves, Björn/Olbrich, Sebastian/Pfeiffer, Daniel* (2009): Forschungsmethodik einer Integrationsdisziplin. In: Becker, Jörg/Krcmar, Helmut/Niehaves, Björn (Hrsg.): Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik. Heidelberg: Physica: 1–22.
- Bellahsene, Zohra/Bonifati, Angela/Duchateau, Fabien/Velegrakis, Yannis* (2011): On Evaluating Schema Matching and Mapping. In: Bellahsene, Zohra/Bonifati, Angela/Rahm, Erhard (Hrsg.): Schema Matching and Mapping. Berlin, Heidelberg: Springer: 253–291.
- Berg, Hartmut/Schmitt, Stefan* (2003): Gloablisierung der Wirtschaft und Kooperation. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. Wiesbaden: Gabler: 277–304.



- Bergamaschi, Sonia/Olaru, Marius-Octavian/Sorrentino, Serena/Vincini, Maurizio* (2011): Semi-automatic Discovery of Mappings Between Heterogeneous Data Warehouse Dimensions. In: *International Journal on Information Technology*, 1 (3): 38–46.
- Berger, Stefan* (2009): *FedDW: A model-driven approach for querying federations of autonomous data marts*. Linz: Johannes Kepler Universität.
- Berghammer, Rudolf* (2012): *Ordnungen, Verbände und Relationen mit Anwendungen*, 2. Aufl. Wiesbaden: Springer Fachmedien.
- Bernstein, Abraham/Kaufmann, Esther/Bürki, Christoph/Klein, Mark* (2005): How Similar Is It? Towards Personalized Similarity Measures in Ontologies. In: *Ferstl, Otto K./Sinz, Elmar J./Eckert, Sven/Isselhorst, Tilman* (Hrsg.): *Wirtschaftsinformatik 2005*. Heidelberg: Physica: 1347–1366.
- Berthold, Henrike/Rösch, Philipp/Zöller, Stefan/Wortmann, Felix/Carenini, Alessio/Campbell, Stuart/Bisson, Pascal/Strohmaier, Frank* (2010): An architecture for ad-hoc and collaborative business intelligence. In: *Daniel, Florian* (Hrsg.): *Proceedings of the 2010 EDBT/ICDT Workshops*. New York, NY: ACM: 1–6.
- Biethahn, Jörg/Lackner, Andreas/Range, Michael* (2004): *Optimierung und Simulation*. München: Oldenbourg.
- Billingsley, Patrick* (1995): *Probability and measure*, 3. Aufl. New York, NY: Wiley.
- Bitterer, Andreas* (2011): Hype Cycle for Business Intelligence, 2011, G00216086. URL: <https://www.gartner.com/doc/1766215/hype-cycle-business-intelligence->, Abruf am 28.11.2014.
- Bitterer, Andreas* (2012): Hype Cycle for Business Intelligence, 2012, G00235187. URL: <https://www.gartner.com/doc/2118915/hype-cycle-business-intelligence->, Abruf am 28.11.2014.
- Bloom, Nicholas/Garicano, Luis/Sadun, Raffaella/van Reenen, John* (2014): The Distinct Effects of Information Technology and Communication Technology on Firm Organization. In: *Management Science*, 60 (12): 2859–2885.
- Böhringer, Martin/Gluchowski, Peter/Kurze, Christian/Schieder, Christian* (2010): A Business Intelligence Perspective on the Future Internet. In: *Association for Information Systems* (Hrsg.): *Proceedings of the 16th American Conference on Information Systems (AMCIS)*. Red Hook, NY: Curran.
- Bonifati, Angela/Chang, Elaine/Ho, Terence/Lakshmanan, Laks V. S./Pottinger, Rachel/Chung, Yonglk* (2010): Schema mapping and query translation in heterogeneous P2P XML databases. In: *The VLDB Journal*, 19: 231–256.
- Borgatti, Stephen P./Everett, Martin G.* (2006): A Graph-theoretic perspective on centrality. In: *Social Networks*, 28 (4): 466–484.
- Bortz, Jürgen/Schuster, Christof* (2010): *Statistik für Human- und Sozialwissenschaftler*, 7. Aufl. Berlin, Heidelberg: Springer.
- Bossel, Hartmut* (1994): *Modellbildung und Simulation*, 2. Aufl. Braunschweig: Vieweg.
- Boyd, Danah M./Ellison, Nicole B.* (2007): Social Network Sites. In: *Journal of Computer-Mediated Communication*, 13 (1): 210–230.



- Brocke, Jan vom* (2003): Referenzmodellierung. Berlin: Logos.
- Brocke, Jan vom/Simons, Alexander/Niehaves, Björn/Riemer, Kai/Plattfaut, Ralf/Cleven, Anne* (2009): Reconstructing the giant: On the importance of rigour in documenting the literature search process. In: Newell, Susan/Whitley, Edgar A./Pouloudi, Nancy/Wareham, Jonathan/Mathiassen, Lars (Hrsg.): Proceedings of the 17th European Conference on Information Systems (ECIS). Verona: Università degli Studi di Verona: 2206–2217.
- Brunner, C. C.* (1998): The New Superintendency Supports an Innovation: Collaborative Decision Making. In: Contemporary Education, 69 (2): 79–82.
- Bühler, Stefan/Jäger, Franz* (2005): Kooperation: Erklärungsperspektive der Industrieökonomik. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. 2. Aufl. Wiesbaden: Gabler: 95–120.
- Bundesamt für Sicherheit in der Informationstechnik* (2008): BSI-Standard 100-1: Managementsysteme für Informationssicherheit (ISMS). URL: [https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard\\_1001\\_pdf.pdf?\\_\\_blob=publicationFile](https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Publikationen/ITGrundschutzstandards/standard_1001_pdf.pdf?__blob=publicationFile), Abruf am 13.11.2014.
- Bundesamt für Sicherheit in der Informationstechnik* (2012): Eckpunktepapier Sicherheitsempfehlungen für Cloud Computing Anbieter. URL: <https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/Mindestanforderungen/Eckpunktepapier-Sicherheitsempfehlungen-CloudComputing-Anbieter.pdf>, Abruf am 27.10.2014.
- Bungartz, Hans-Joachim/Buchholz, Martin/Pflüger, Dirk/Zimmer, Stefan* (2009): Modellbildung und Simulation. Berlin, Heidelberg: Springer.
- Burmester, Lars* (2011): Adaptive Business-Intelligence-Systeme. Wiesbaden: Vieweg+Teubner.
- Busemann, Katrin/Gscheidle, Christoph* (2012): Web 2.0: Habitualisierung der Social Communitys. In: MEDIA PERSPEKTIVEN (7-8): 380–390.
- Busse, Johannes/Humm, Bernhard/Lübbert, Christoph/Moelter, Frank/Reibold, Anatol/Rewald, Matthias/Schlüter, Veronika/Seiler, Bernhard/Tegtmeier, Erwin/Zeh, Thomas* (2014): Was bedeutet eigentlich Ontologie? In: Informatik Spektrum, 37 (4): 286–297.
- Cabibbo, Luca/Torlone, Riccardo* (1998): A Logical Approach to Multidimensional Databases. In: Schek, Hans-Jörg/Alonso, Gustavo/Saltor, Felix/Ramos, Isidro (Hrsg.): Advances in Database Technology EDBT'98. Berlin, Heidelberg: Springer: 183–197.
- Cabibbo, Luca/Torlone, Riccardo* (2004): On the Integration of Autonomous Data Marts. In: Hatzopoulos Michael/Manolopoulos Yannis (Hrsg.): Proceedings of the 16th International Conference on Scientific and Statistical Database Management (SSDBM 2004). Los Alamitos, CA: IEEE: 223–231.
- Calì, Andrea/Calvanese, Diego/Giacomo, Giuseppe de/Lenzerini, Maurizio* (2003): On the Expressive Power of Data Integration Systems. In: Spaccapietra, Stefano/March, Salvatore T./Kambayashi, Yahiko (Hrsg.): Conceptual Modeling — ER 2002. Berlin, Heidelberg: Springer: 338–350.
- Camarinha-Matos, Luis M./Afsarmanesh, Hamideh/Galeano, Nathalie/Molina, Arturo* (2009): Collaborative networked organizations – Concepts and

- practice in manufacturing enterprises. In: Collaborative e-Work Networks in Industrial Engineering, 57 (1): 46–60.
- Cannataro, Mario/Cluet, Sophie/Tradigo, Giuseppe/Veltri, Pierangelo/Vodislav, Dan* (2006): Using Views to Query XML Documents. In: Rivero, Laura C./Doorn, Jorge H./Ferraggine, Viviana E. (Hrsg.): Encyclopedia of database technologies and applications. Hershey, PA: Idea Group Reference: 729–735.
- Cate, Balder ten/Kolaitis, Phokion G.* (2010): Structural characterizations of schema-mapping languages. In: Communications of the ACM, 53 (1): 101–110.
- Chamoni, Peter/Beeckmann, Frank/Bley, Tanja* (2010): Ausgewählte Verfahren des Data Mining. In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 329–356.
- Chamoni, Peter/Gluchowski, Peter* (2000): On-Line Analytical Processing (OLAP). In: Mucksch, Harry/Behme, Wolfgang (Hrsg.): Das Data Warehouse-Konzept. 4. Aufl. Wiesbaden: Gabler: 333–376.
- Chamoni, Peter/Gluchowski, Peter* (2004): Integrationstrends bei Business-Intelligence-Systemen. In: WIRTSCHAFTSINFORMATIK, 46 (2): 119–128.
- Chamoni, Peter/Gluchowski, Peter* (2010): Analytische Informationssysteme - Einordnung und Überblick. In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 3–16.
- Chang, Elizabeth/Dillon, Tharam S./Hussain, Farookh K.* (2006): Trust and reputation for service oriented environments. Chichester: Wiley.
- Chaudhuri, Surajit/Dayal, Umeshwar/Narasayya, Vivek* (2011): An overview of business intelligence technology. In: Communications of the ACM, 54 (8): 88–98.
- Chehade, Rafic Hage/El Ghor, Hussein/Nachouki, Gilles/Fliti, Tamim* (2009): Semantic Routing in Peer-to-Peer Systems: Techniques and Categories. In: Bouhlef, Mohamed S./Puech, William (Hrsg.): Proceedings of the 5th International Conference: Sciences of Electronics, Technologies of Information and Telecommunication (SETIT 2009). Hammamet, Tunisia: IEEE: 1–12.
- Chen, Hsinchun/Chiang, Roger H./Storey, Veda C.* (2012): Business intelligence and analytics. In: MIS Quarterly, 36 (4): 1165–1188.
- Chen, Jianguo/Xu, Xiang/Bruda, S. D.* (2010): Combining data trust in reputation systems to boost P2P security. In: IEEE (Hrsg.): Proceedings of the 2nd International Conference on Future Computer and Communication (ICFCC). Piscataway, NJ: IEEE: 194–199.
- Chen, Min/Mao, Shiwen/Liu, Yunhao* (2014): Big Data: A Survey. In: Mobile Networks and Applications, 19 (2): 171–209.
- Chen, Peter P.-S.* (1976): The entity-relationship model - toward a unified view of data. In: ACM Transactions on Database Systems, 1 (1): 9–36.

- Chen, Qinghua/Shi, Dinghua* (2004): The modeling of scale-free networks. In: *Physica A: Statistical Mechanics and its Applications*, 335 (1–2): 240–248.
- Chen, Xuhui/Li, Hongyong/Ke, Ming* (2012): Study on Complex Network Stability Based on the Structure Parameters. In: Association for Information Systems (Hrsg.): *Proceedings of the 2012 International Conference on Information Systems (ICIS)*. Red Hook, NY: Curran: 521–526.
- Cirani, Simone/Melegari, Lorenzo/Veltri, Luca* (2010): Peer-to-peer technologies applied to data warehouses. In: IEEE (Hrsg.): *GLOBECOM Workshops (GC Wkshps)*. Piscataway, NJ: IEEE: 1952–1956.
- Clark, John/Holton, Derek A.* (1994): *Graphentheorie*. Heidelberg, Berlin, Oxford: Spektrum.
- Cleff, Thomas* (2015): *Deskriptive Statistik und Explorative Datenanalyse*. Wiesbaden: Gabler.
- Codd, Edgar F.* (1979): Extending the database relational model to capture more meaning. In: *ACM Transactions on Database Systems*, 4 (4): 397–434.
- Codd, Edgar F./Codd, Sally B./Salley, Clynn T.* (1993): *Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate*. O.O.: E.F. Codd Associates.
- Coenenberg, Adolf G./Salfeld, Rainer* (2007): *Wertorientierte Unternehmensführung*, 2. Aufl. Stuttgart: Schäffer-Poeschel.
- Cohen, Sara/Nutt, Werner/Sagiv, Yehoshua* (2006): Rewriting queries with arbitrary aggregation functions using views. In: *ACM Transactions on Database Systems*, 31 (2): 672–715.
- Collins, Heidi* (2001): *Corporate portals: revolutionizing information access to increase productivity and drive the bottom line*. New York, NY: Amacom Books.
- Cooper, Harris* (1988): Organizing knowledge syntheses: A taxonomy of literature reviews. In: *Knowledge in Society*, 1 (1): 104–126.
- Cooper, Harris/Hedges, Larry V.* (1994): Research Synthesis as a Scientific Enterprise. In: Cooper, Harris/Hedges, Larry V. (Hrsg.): *The Handbook of Research Synthesis*. O.O.: Russel Sage foundation: 3–14.
- Cooper, Martha C./Lambert, Douglas M./Pagh, Janus D.* (1997): Supply Chain Management. In: *The International Journal of Logistics Management*, 8 (1): 1–14.
- Crespo, Arturo/Garcia-Molina, Hector* (2005): Semantic Overlay Networks for P2P Systems. In: Moro, Gianluca/Bergamaschi, Sonia/Aberer, Karl (Hrsg.): *Agents and Peer-to-Peer Computing*. Berlin, Heidelberg: Springer: 1–13.
- Cross, Nigel* (2001): Designerly ways of knowing: design discipline versus design science. In: *Design Issues*, 17 (3): 49–55.
- Cuzzocrea, Alfredo/Bertino, Elisa* (2011): Privacy Preserving OLAP over Distributed XML Data. In: *Journal of Computer and System Sciences*, 77: 965–987.
- Cuzzocrea, Alfredo/Bertino, Elisa* (2014): A Comprehensive Theoretical Framework for Privacy Preserving Distributed OLAP. In: Meersman, Rob-

- ert/Panetto, Hervé/Mishra, Alok/Valencia-García, Rafael/Soares, António/Lucas/Ciuciu, Ioana/Ferri, Fernando/Weichhart, Georg/Moser, Thomas/Bezzi, Michele/Chan, Henry (Hrsg.): *On the Move to Meaningful Internet Systems: OTM 2014 Workshops*. Berlin, Heidelberg: Springer: 117–136.
- Dangalchev, Chavdar* (2004): Generation models for scale-free networks. In: *Physica A: Statistical Mechanics and its Applications*, 338 (3–4): 659–671.
- Dangelmaier, Wilhelm/Laroque, Christoph* (2014): Simulation. URL: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/technologien-methoden/Operations-Research/Simulation>, Abruf am 19.03.2015.
- Das, T. K./Teng, Bing-Sheng* (2000): A Resource-Based Theory of Strategic Alliances. In: *Journal of Management*, 26 (1): 31–61.
- Dayal, Umeshwar/Vennelakanti, Ravigopal/Sharma, Ratnesh/Castellanos, Malu/Hao, Ming/Patel, Chandrakant* (2008): Collaborative Business Intelligence: Enabling Collaborative Decision Making in Enterprises. In: Tari, Zahir/Meersmann, Robert (Hrsg.): *On the Move to Meaningful Internet Systems: OTM 2008*. Berlin, Heidelberg: Springer: 8–25.
- Dayal, Umeshwar/Vennelakanti, Ravigopal/Sharma, Ratnesh K./Gastellanos, Maria G./Hao, Ming C./Patel, Chandrakant/Bellad, Sangamesh S./Gupta, Manish* (2010): Providing Collaborative Business Intelligence. Patentnr. US2010/0325206A1.
- Delveroudis, Yannis/Lekeas, Paraskevas* (2007): Managing Semantic Loss during Query Reformulation in Peer Data Management Systems. In: IEEE (Hrsg.): *Proceedings of the 2007 IEEE International Workshop on Databases for Next Generation Researchers (SWOD '07)*. Washington, DC: IEEE: 61–66.
- Deming, William E.* (2000): *Out of the crisis*. Cambridge, MA et al.: MIT Press.
- DeSanctis, Gerardine/Gallupe, R. B.* (1987): A Foundation for the Study of Group Decision Support Systems. In: *Management Science*, 33 (5): 589–609.
- Devlin, Barry* (1997): *Data warehouse*. Reading, MA et al.: Addison-Wesley.
- Devlin, Barry* (2012): Collaborative Business Intelligence: Socializing Team-Based Decision Making. In: *Business Intelligence Journal*, 17 (3): 9–17.
- Dewitt, David J./Gray, Jim* (1992): Parallel database systems: the future of high performance database systems. In: *Communications of the ACM*, 35 (6): 85–98.
- Diestel, Reinhard* (1996): *Graphentheorie*. Berlin, Heidelberg: Springer.
- Dinter, Barbara/Lorenz, Anja* (2012): Social Business Intelligence: a Literature Review and Research Agenda. In: Association for Information Systems (Hrsg.): *Proceedings of the 2012 International Conference on Information Systems (ICIS)*. Red Hook, NY: Curran: 1–12.
- Doan, AnHai/Domingos, Pedro/Halevy, Alon* (2003): Learning to Match the Schemas of Data Sources: A Multistrategy Approach. In: *Machine Learning*, 50 (3): 279–301.
- Doan, AnHai/Madhavan, Jayant/Domingos, Pedro/Halevy, Alon* (2004): Ontology Matching: A Machine Learning Approach. In: Staab, Steffen/Studer,

- Rudi (Hrsg.): Handbook on Ontologies. Berlin, Heidelberg: Springer: 385–403.
- Doka, Katerina/Tsoumakos, Dimitrios/Koziris, Nectarios* (2011): Brown Dwarf: A fully-distributed, fault-tolerant data warehousing system. In: Journal of Parallel and Distributed Computing, 71 (11): 1434–1446.
- Domschke, Wolfgang/Drexl, Andreas* (2011): Einführung in Operations Research, 8. Aufl. Berlin, Heidelberg: Springer.
- Doz, Yves* (1992): Empirische Relevanz von Strategischen Allianzen in Europa. In: Bronder, Christoph (Hrsg.): Wegweiser für strategische Allianzen. Frankfurt am Main, Wiesbaden: Frankfurter Allgemeine, Gabler: 47–62.
- Dresner, Howard* (2013): Dresner Report on Collaborative Business Intelligence Trends Reveals Where Collaboration is Most Valued. URL: <http://sandhill.com/article/dresner-report-on-collaborative-business-intelligence-trends-reveals-where-collaboration-is-most-valued/>, Abruf am 28.11.2014.
- Duchateau, Fabien/Bellahsene, Zohra* (2010): Measuring the Quality of an Integrated Schema. In: Parsons, Jeffrey/Saeki, Motoshi/Shoval, Peretz/Woo, Carson/Wand, Yair (Hrsg.): Conceptual Modeling – ER 2010. Berlin, Heidelberg: Springer: 261–273.
- Dudek, Gregor* (2009): Collaborative Planning. In: Zeitschrift für Planung & Unternehmenssteuerung, 20 (2): 239–246.
- Duden* (2015a): Fairness. URL: <http://www.duden.de/rechtschreibung/Fairness>, Abruf am 11.05.2015.
- Duden* (2015b): Kollaboration. URL: <http://www.duden.de/rechtschreibung/Kollaboration>, Abruf am 11.05.2015.
- Duden* (2015c): Kooperation. URL: <http://www.duden.de/rechtschreibung/Kooperation>, Abruf am 11.05.2015.
- Duden* (2015d): Kriterium. URL: <http://www.duden.de/rechtschreibung/Kriterium>, Abruf am 11.05.2015.
- Duden* (2015e): Merkmal. URL: <http://www.duden.de/rechtschreibung/Merkmal>, Abruf am 11.05.2015.
- Duden* (2015f): Qualität. URL: <http://www.duden.de/rechtschreibung/Qualitaet>, Abruf am 11.05.2015.
- Düsing, Roland* (2010): Knowledge Discovery in Databases. In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 281–306.
- Dyche, Jill* (2012): End the Chaos: A Better Approach to Business Intelligence. URL: <https://hbr.org/2012/05/end-the-chaos-a-better-approac>, Abruf am 17.03.2015.
- Eckey, Hans-Friedrich/Kosfeld, Reinhold/Rengers, Martina* (2002): Multivariate Statistik. Wiesbaden: Gabler.
- Eckhardt, Alan/Skopal, Tomáš/Vojtáš, Peter* (2009): On Fuzzy vs. Metric Similarity Search in Complex Databases. In: Andreasen, Troels/Yager, Ronald



- R./Bulskov, Henrik/Christiansen, Henning/Larsen, Henrik L. (Hrsg.): Flexible Query Answering Systems. Berlin, Heidelberg: Springer: 64–75.
- Eckstein, Jonathan* (2013): Parallel Computing. In: Gass, Saul I./Fu, Michael C. (Hrsg.): Encyclopedia of Operations Research and Management Science. 3. Aufl. Boston, MA: Springer US: 1103–1107.
- Edmundson, Anne/Holtkamp, Brian/Rivera, Emanuel/Finifter, Matthew/Mettler, Adrian/Wagner, David* (2013): An Empirical Study on the Effectiveness of Security Code Review. In: Jürjens, Jan/Livshits, Benjamin/Scandariato, Riccardo (Hrsg.): Engineering Secure Software and Systems: Springer Berlin Heidelberg: 197–212.
- Eisenführ, Franz/Weber, Martin* (1999): Rationales Entscheiden, 3. Aufl. Berlin, Heidelberg: Springer.
- Ellis, Donald G./Fisher, B. A.* (1994): Small group decision making, 4. Aufl. New York et al.: McGraw-Hill.
- EMA Inc.* (2012): The Role of Collaboration in Business Intelligence. URL: <http://www.enterprisemanagement.com/research/asset-free.php/2410/pre/The-Role-of-Collaboration-in-Business-Intelligence-pre>, Abruf am 11.12.2013.
- Engelhard, Johann/Altmann, Jörn* (2015): Joint Venture. Gabler Wirtschaftslexikon. URL: <http://wirtschaftslexikon.gabler.de/Archiv/968/joint-venture-v10.html>, Abruf am 11.05.2015.
- English, Larry P.* (2009): Information quality applied. Indianapolis, IN: Wiley.
- Erdmann, Georg/Popp, Harald* (2006): Betriebswirtschaft, 4. Aufl. Karlsruhe: Versicherungswirtschaft.
- Fagin, Ronald* (1999): Combining Fuzzy Information from Multiple Systems. In: Journal of Computer and System Sciences, 58 (1): 83–99.
- Fayyad, Usama M./Piatetsky-Shapiro, Gregory/Smyth, Padhraic* (1996a): Advances in knowledge discovery and data mining. In: Fayyad, Usama M./Piatetsky-Shapiro, Gregory/Smyth, Padhraic/Uthurusamy, Ramasamy (Hrsg.): Advances in knowledge discovery and data mining. Menlo Park et al.: AAAI Press et al.: 1–34.
- Fayyad, Usama M./Piatetsky-Shapiro, Gregory/Smyth, Padhraic* (1996b): From Data Mining to Knowledge Discovery in Databases. In: AI Magazine, 1996 (FALL): 37–54.
- Fellbaum, Christiane* (2005): WordNet and wordnets. In: Brown, Keith (Hrsg.): Encyclopedia of Language and Linguistics. 2. Aufl. Amsterdam et al.: Elsevier: 665–670.
- Fettke, Peter* (2006): State-of-the-Art des State-of-the-Art. In: WIRTSCHAFTSINFORMATIK, 48 (4): 257–266.
- Fettke, Peter/Brocke, Jan vom* (2013): Referenzmodell. URL: <http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/is-management/Systementwicklung/Softwarearchitektur/Wiederverwendung-von-Softwarebausteinen/Referenzmodell/index.html>, Abruf am 11.05.2015.

- Fettke, Peter/Loos, Peter* (2003): Classification of reference models: a methodology and its application. In: *Information Systems and e-Business Management*, 1 (1): 35–53.
- Fettke, Peter/Loos, Peter* (2004): Entwicklung eines Bezugsrahmens zur Evaluierung von Referenzmodellen – Langfassung eines Beitrages. Mainz: Johannes Gutenberg-Universität Mainz, Research Group Information Systems & Management.
- Fischer, Joachim/Herold, Werner/Dangelmaier, Wilhelm/Nastansky, Ludwig/Suhl, Leena* (2008): Bausteine der Wirtschaftsinformatik, 4. Aufl. Berlin: E. Schmidt.
- Fontanari, Martin* (1996): Kooperationsgestaltungsprozesse in Theorie und Praxis. Berlin: Duncker & Humblot.
- Foster, Ian* (2002): What is the Grid? A Three Point Checklist. URL: <http://dlib.cs.odu.edu/WhatIsTheGrid.pdf>, Abruf am 21.04.2105.
- Foster, Ian/Kesselman, Carl* (2004): The grid - Blueprint for a new computing infrastructure, 2. Aufl. Amsterdam: Morgan Kaufmann.
- Frank, Ulrich* (2006): Towards a pluralistic conception of research methods in information systems research. Essen: Institut für Informatik und Wirtschaftsinformatik (ICB), Universität Duisburg-Essen.
- Franz, Timothy M.* (2012): Group Dynamics and Team Interventions. Hoboken: John Wiley & Sons.
- Freire, Crishane/Souza, Bruno F. F./Souza, Damires/Batista, Maria C. M./Salgado, Ana C.* (2012): Towards an information quality approach to enhance query routing processes. In: Pardede, Eric (Hrsg.): Proceedings of the 14th International Conference on Information Integration and Web-based Applications & Services (IIWAS '12). New York, NY: ACM: 311–314.
- Freire, Crishane/Souza, Bruno F. F./Souza, Damires/Batista, Maria C. M./Salgado, Ana C.* (2013): Semantic measures as information quality criteria for query routing processes. In: *International Journal of Business Intelligence and Data Mining*, 8 (2): 167–183.
- Freund, Yoav/Mansour, Yishay/Schapire, Robert E.* (2001): Why Averaging Classifiers Can Protect Against Overfitting. In: Jaakkola, Tommi S./Richardson, Thomas (Hrsg.): Artificial intelligence and statistics 2001. San Francisco, CA: Morgan Kaufmann: 1–8.
- Furrer, Frank J.* (2014): Eine kurze Geschichte der Ontologie. In: *Informatik Spektrum*, 37 (4): 308–317.
- Furtado, Pedro* (2009): A Survey of Parallel and Distributed Data Warehouses. In: *International Journal of Data Warehousing and Mining*, 5 (2): 57–77.
- Gansor, Tom/Totok, Andreas/Stock, Steffen* (2010): Von der Strategie zum Business Intelligence Competency Center (BICC). München: Hanser.
- Gantz, John/Reinsel, David* (2011): IDC iView: Extracting Value from Chaos. URL: <http://www.emc.com/collateral/analyst-reports/idc-extracting-value-from-chaos-ar.pdf>, Abruf am 30.10.2014.

- Gartner Inc.* (2013): Hype Cycles. URL: <http://www.gartner.com/technology/research/methodologies/hype-cycle.jsp>, Abruf am 17.12.2013.
- Geist, Fabian/Kluin, Torsten/Ritz, Harald* (2013): Self-Service Business Intelligence (SSBI). In: Barton, Thomas/Erdlenbruch, Burkhard/Guckert, Michael/Herrmann, Frank/Müller, Christian/Ritz, Harald (Hrsg.): Herausforderungen an die Wirtschaftsinformatik: Integration und Konnexion. Berlin: News & Media: 47–58.
- Gericke, Anke/Winter, Robert* (2009): Entwicklung eines Bezugsrahmens für Konstruktionsforschung und Artefaktkonstruktion in der gestaltungsorientierten Wirtschaftsinformatik. In: Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik: 195–210.
- Gerlach, Max/Cleophas, Catherine/Kliwer, Natalia* (2013): Codeshare-Allianzen bei Fluggesellschaften. In: WIRTSCHAFTSINFORMATIK, 55 (3): 147–158.
- Ghobadi, Shahla/Clegg, Stewart* (2015): “These days will never be forgotten ...”: A critical mass approach to online activism. In: Information and Organization, 25 (1): 52–71.
- Gibbs, Richard/Humphries, Andrew* (2009): Strategic alliances & marketing partnerships. London et al.: Kogan Page.
- Gilbert, Nigel/Troitzsch, Klaus G.* (2005): Simulation for the social scientist, 2. Aufl. Maidenhead, Berkshire et al.: Open University Press et al.
- Gill, T. G./Hevner, Alan R.* (2013): A Fitness-Utility Model for Design Science Research. In: ACM Transaction on Management Information Systems, 4 (2): Article No. 5.
- Gilles, Daniel* (2007): Modelling, Implementing and Exploring Agent-based Models: An Example. In: Phan, Denis/Amblard, Frédéric (Hrsg.): Agent based modelling and simulation in the social and human sciences. Oxford: Bardwell Press: 257–272.
- Gleeson, James P./Cahalane, Diarmuid J.* (2007): Seed size strongly affects cascades on random networks. In: Physical Review E, 75 (5): 1–4.
- Gluchowski, Peter* (2001): Business Intelligence - Konzepte, Technologien und Einsatzbereiche. In: HMD Praxis der Wirtschaftsinformatik (222): 5–15.
- Gluchowski, Peter/Chamoni, Peter* (2010): Entwicklungslinien und Architekturkonzepte des OLAP. In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 197–228.
- Gluchowski, Peter/Dittmar, Carsten/Gabriel, Roland* (2008): Management Support Systeme und Business Intelligence, 2. Aufl. Berlin, Heidelberg: Springer.
- Goldreich, Oded/Goldwasser, Shafi/Micali, Silvio* (1986): How to construct random functions. In: Journal of the ACM, 33 (4): 792–807.
- Golfarelli, Matteo* (2009): Open Source BI Platforms: A Functional and Architectural Comparison. In: Hutchison, David/Kanade, Takeo/Kittler, Josef/Kleinberg, Jon M./Mattern, Friedemann/Mitchell, John C./Naor, Moni/Nierstrasz, Oscar/Pandu Rangan, C./Steffen, Bernhard/Sudan,



- Madhu/Terzopoulos, Demetri/Tygar, Doug/Vardi, Moshe Y./Weikum, Gerhard/Pedersen, Torben B./Mohania, Mukesh K./Tjoa, A. Min (Hrsg.): Data Warehousing and Knowledge Discovery. Berlin, Heidelberg: Springer: 287–297.
- Golfarelli, Matteo/Mandreoli, Federica/Penzo, Wilma/Rizzi, Stefano/Turricchia, Elisa* (2010): Towards OLAP query reformulation in peer-to-peer data warehousing. In: Song, Il-Yeol/Ordonez, Carlos (Hrsg.): Proceedings of the 13th ACM International Workshop on Data Warehousing and OLAP (DOLAP). New York, NY: ACM: 37–44.
- Golfarelli, Matteo/Mandreoli, Federica/Penzo, Wilma/Rizzi, Stefano/Turricchia, Elisa* (2011): BIN: Business intelligence networks. In: Zorrilla, Marta E./Mazón, Jose-Norberto/Ferrández, Óscar/Garrigós, Irene/Daniel, Florian/Trujillo, Juan (Hrsg.): Business Intelligence Applications and the Web. Hershey, PA: IGI Global: 244–265.
- Golfarelli, Matteo/Mandreoli, Federica/Penzo, Wilma/Rizzi, Stefano/Turricchia, Elisa* (2012a): A Query Reformulation Framework for P2P OLAP. In: Nicola Ferro/Letizia Tanca (Hrsg.): Proceedings of the 20th Italian Symposium on Advanced Database Systems (SEBD). Padova: Edizioni Libreria Progetto: 147–154.
- Golfarelli, Matteo/Mandreoli, Federica/Penzo, Wilma/Rizzi, Stefano/Turricchia, Elisa* (2012b): OLAP query reformulation in peer-to-peer data warehousing. In: Information Systems, 37 (5): 393–411.
- Göltenboth, Markus* (1998): Global Sourcing und Kooperationen als Alternativen zur vertikalen Integration. Frankfurt am Main et al.: Lang.
- Gong, Shang-Fu/Zhu, Jian-Lei* (2012): A Survey of Reputation and Trust Mechanism in Peer-to-Peer Network. In: IEEE (Hrsg.): Proceedings of the 2012 International Conference on Industrial Control and Electronics Engineering (ICICEE). Los Alamitos, CA: IEEE.
- Gould, Ronald J.* (2014): Hamilton Graphs. In: Gross, Jonathan L./Yellen, Jay/Zhang, Ping (Hrsg.): Handbook of graph theory. 2. Aufl. Boca Raton, FL: CRC Press: 314–335.
- Gray, Paul* (1987): Group decision support systems. In: Decision Support Systems, 3 (3): 233–242.
- Gregor, Shirley/Hevner, Alan R.* (2013): Positioning and Presenting Design Science Research for Maximum Impact. In: MIS Quarterly, 37 (2): 337–355.
- Gross, Jonathan L./Yellen, Jay* (2006): Graph theory and its applications, 2. Aufl. Boca Raton, FL: Chapman & Hall/CRC.
- Gross, Jonathan L./Yellen, Jay* (2014): Fundamentals of Graph Theory. In: Gross, Jonathan L./Yellen, Jay/Zhang, Ping (Hrsg.): Handbook of graph theory. 2. Aufl. Boca Raton, FL: CRC Press: 2–20.
- Gross, Jonathan L./Yellen, Jay/Zhang, Ping* (Hrsg.) (2014): Handbook of graph theory, 2. Aufl. Boca Raton, FL: CRC Press.
- Grubestic, Tony H./Matisziw, Timothy C./Murray, Alan T./Snediker, Diane* (2008): Comparative Approaches for Assessing Network Vulnerability. In: International Regional Science Review, 31 (1): 88–112.

- Gulati, Ranjay* (1995): The Implications of Repeated Ties for Contractual Choice in Alliances. In: *The Academy of Management Journal*, 38 (1): 85–112.
- Gurjar, Yuvraj S./Rathore, Vijay S.* (2013): Cloud Business Intelligence – Is What Business Need Today. In: *International Journal of Recent Technology and Engineering (IJRTE)*, 1 (6): 81–86.
- Hägele, Markus* (2015): HR Analytics bei F. Hoffmann-La Roche Ltd. In: *Strohmeier, Stefan/Piazza, Franca* (Hrsg.): *Human Resource Intelligence und Analytics*. Wiesbaden: Springer Fachmedien: 305–335.
- Hagerty, John/Sallam, Rita L./Richardson, James* (2012): Magic quadrant for Business Intelligence Platforms, G00225500. URL: <https://www.gartner.com/doc/1915014/magic-quadrant-business-intelligence-platforms>, Abruf am 28.11.2014.
- Hahn, Dietger/Hungenberg, Harald* (2001): *PuK - Planung und Kontrolle*, 6. Aufl. Wiesbaden: Gabler.
- Hahne, Michael* (2010): Mehrdimensionale Datenmodellierung für analyseorientierte Informationssysteme. In: *Chamoni, Peter/Gluchowski, Peter* (Hrsg.): *Analytische Informationssysteme*. 4. Aufl. Berlin, Heidelberg: Springer: 229–258.
- Hahne, Michael/Kurze, Christian* (2012): OLAP-Modellierung mit Hilfe der ADAPT-Notation. In: *BI Spektrum*, 7 (1): 23–27.
- Halang, Wolfgang A./Konakovsky, Rudolf M.* (2013): *Sicherheitsgerichtete Echtzeitsysteme*, 2. Aufl. Berlin, Heidelberg: Springer.
- Halevy, Alon Y.* (2001): Answering queries using views: A survey. In: *The VLDB Journal*, 10 (4): 270–294.
- Halevy, Alon Y./Ives, Zachary G./Madhavan, Jayant/Mork, Peter/Suciu, Dan/Tatarinov, Igor* (2004): The Piazza peer data management system. In: *IEEE Transactions on Knowledge and Data Engineering*, 16 (7): 787–798.
- Halevy, Alon Y./Ives, Zachary G./Suciu, Dan/Tatarinov, Igor* (2003): Schema mediation in peer data management systems. In: *Dayal, Umeshwar/Ramamritham, Krithi/Vijayaraman, T. M.* (Hrsg.): *Proceedings of the 19th International Conference on Data Engineering*. Piscataway, NJ: IEEE: 505–516.
- Hammer, Richard M.* (2011): *Planung und Führung*, 8. Aufl. München: Oldenbourg.
- Hammergren, Tom/Simon, Alan R.* (2009): *Data warehousing for dummies*, 2. Aufl. Hoboken, NJ: Wiley.
- Hammond, Mark* (1999): SAS analyzes disparate data. In: *PC Week*, 16 (8): 117.
- Hansen, Hans R./Neumann, Gustaf* (2009): *Wirtschaftsinformatik 1*, 10. Aufl. Stuttgart: Lucius & Lucius.
- Hartmann, Stefan* (2008): Überwindung semantischer Heterogenität bei multiplen Data-Warehouse-Systemen. In: *Schriften aus der Fakultät Wirtschaftsinformatik und Angewandte Informatik der Otto-Friedrich-Universität Bamberg*. Bamberg: University of Bamberg Press: 1–251.

- Hauschild, Wolfgang/Wallacher, Ludwig* (2004): Ad-hoc-Befragung über Unternehmenskooperationen. In: Statistisches Bundesamt (Hrsg.): Wirtschaft und Statistik. Wiesbaden: Statistisches Bundesamt: 1009–1016.
- Heinrich, Lutz J.* (2005): Forschungsmethodik einer Integrationsdisziplin: Ein Beitrag zur Geschichte der Wirtschaftsinformatik. In: NTM International Journal of History and Ethics of Natural Sciences, Technology and Medicine, 13 (2): 104–117.
- Heinrich, Lutz J./Heinzl, Armin/Roithmayr, Friedrich* (2007): Wirtschaftsinformatik, 3. Aufl. München: Oldenbourg.
- Herzwurm, Georg/Stelzer, Dirk* (2008): Wirtschaftsinformatik versus Information Systems - Eine Gegenüberstellung. Ilmenau: Techn. Univ. Inst. für Wirtschaftsinformatik.
- Heßler, Armin/Mosebach, Petra* (2013): Strategie und Marketing im Web 2.0. Wiesbaden: Gabler.
- Hettler, Uwe* (2010): Social-Media-Marketing. München: Oldenbourg.
- Hevner, Alan R./March, Salvatore T./Park, Jinsoo/Ram, Sudha* (2004): Design Science in Information Systems Research. In: MIS Quarterly, 28 (1): 75–105.
- Heymann, Sébastien* (2014): Gephi. In: Alhajj, Reda/Rockne, Jon (Hrsg.): Encyclopedia of Social Network Analysis and Mining. New York: Springer: 612–625.
- Hirschmann, Petra* (1998): Kooperative Gestaltung unternehmenübergreifender Geschäftsprozesse. Wiesbaden: Gabler.
- Hoffmann, Dirk W.* (2013): Software-Qualität, 2. Aufl. Berlin: Springer Vieweg.
- Holl, Albert/Auerochs, Robert* (2004): Analogisches Denken als Erkenntnisstrategie zur Modellbildung in der Wirtschaftsinformatik. In: Frank, Ulrich (Hrsg.): Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Wiesbaden: DUV: 367–389.
- Horakova, Marketa/Skalska, Hana* (2013): Business Intelligence and Implementation in a Small Enterprise. In: Journal of Systems Integration, 4 (2): 50–61.
- Hosack, Bryan/Hall, Dianne/Paradice, David B./Courtney, James F.* (2012): A Look Toward the Future: Decision Support Systems Research is Alive and Well. In: Journal of the Association for Information Systems, 13 (5): 315–340.
- Howard, Mindy S.* (1994): Quality of group decision support systems. Eindhoven: Eindhoven University of Technology.
- Huang, Yun/Shen, Cuihua/William, Dmitri/Contractor, Noshir* (2009): Virtually There: Exploring Proximity and Homophily in a Virtual World. In: IEEE (Hrsg.): 2009 International Conference on Computational Science and Engineering. Piscataway, NJ: IEEE: 354–359.
- Huch, Burkhard/Behme, Wolfgang/Ohlendorf, Thomas* (2004): Rechnungswesenorientiertes Controlling, 4. Aufl. Heidelberg: Physica.
- Hummeltenberg, Wilhelm* (2010): Vom Content Managment zum Enterprise Decision Management - eine Chronologie der Business Intelligence-Systeme.

- In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 17–36.
- Hurtado, Carlos A./Mendelzon, Alberto O.* (2002): OLAP dimension constraints. In: Abiteboul, Serge (Hrsg.): Proceedings of the twenty-first ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems. Madison, WI: ACM: 169–179.
- Hutschenreuter, Thomas* (2009): Allgemeine Betriebswirtschaftslehre, 3. Aufl. Wiesbaden: Gabler.
- Ibragimov, Dilshod/Hose, Katja/Pedersen, Torben B./Zimányi, Esteban* (2014): Towards Exploratory OLAP over Linked Open Data - A Case Study. In: Proceedings of the 8th International Workshop on Business Intelligence for the Real-Time Enterprise (BIRTE 2014). Berlin: Springer: 1–18.
- Ihrig, Falk* (1991): Strategische Allianzen. In: Wirtschaftswissenschaftliches Studium, 20 (1): 29–31.
- Imhoff, Claudia/White, Colin* (2011): Self-Service Business Intelligence: Empowering Users to Generate Insights. URL: <http://spots.aug.edu/tschultz/classes/201402Minf3650/Content/Self-Service-BI-Empowerment.pdf>, Abruf am 12.03.2014.
- Inmon, William H.* (2005): Building the Data Warehouse, 4. Aufl. Indianapolis: Wiley.
- Inmon, William H./Strauss, Derek/Neushloss, Genia* (2008): DW 2.0. Amsterdam et al.: Elsevier/Morgan Kaufmann.
- International Telecommunication Union* (1994): Information technology – Open Systems Interconnection – Basic Reference Model: The basic model. URL: <http://www.itu.int/ITU-T/recommendations/rec.aspx?rec=2820>, Abruf am 19.11.2014.
- Irle, Christof* (2011): Rationalität von Make-or-buy-Entscheidungen in der Produktion. Wiesbaden: Gabler.
- ISO 15704* (2000): Industrial automation Systems - Requirements for enterprise-reference architectures and methodologies. Genf: International Organization for Standardization.
- Jackson, Matthew O.* (2008): Social and economic networks. Princeton, N.J., Woodstock: Princeton University Press.
- Jamakovic, Almerima/Uhlig, Steve* (2007): Influence of the network structure on robustness. In: IEEE (Hrsg.): 15th IEEE International Conference on Networks, 2007 (ICON 2007). Adelaide, Australia: IEEE: 278–283.
- Ji, Qiu/Haase, Peter/Qi, Guilin* (2011): Combination of Similarity Measures in Ontology Matching Using the OWA Operator. In: Yager, RonaldR/Kacprzyk, Janusz/Beliakov, Gleb (Hrsg.): Recent Developments in the Ordered Weighted Averaging Operators: Theory and Practice. Berlin, Heidelberg: Springer: 281–295.
- Juan, Angel A./Faulin, Javier/Jorba, Josep/Caceres, Jose/Marquès, Joan M.* (2013): Using parallel & distributed computing for real-time solving of vehicle routing problems with stochastic demands. In: Annals of Operations Research, 207 (1): 43–65.

- Kagermann, Henning* (2014): Chancen von Industrie 4.0 nutzen. In: Bauernhansl, Thomas/Hompel, Michael ten/Vogel-Heuser, Birgit (Hrsg.): Industrie 4.0 in Produktion, Automatisierung und Logistik. Wiesbaden: Springer Fachmedien: 603-614.
- Kahn, Beverly K./Strong, Diane M./Wang, Richard Y.* (2002): Information Quality Benchmarks: Product and Service Performance. In: Communications of the ACM, 45 (4): 184–192.
- Kalnis, Panos/Ng, Wee S./Ooi, Beng C./Papadias, Dimitris/Tan, Kian-Lee* (2002): An adaptive peer-to-peer network for distributed caching of OLAP results. In: Franklin, Michael J./Moon, Bongki/Ailamaki, Anastassia (Hrsg.): Proceedings of the 2002 ACM SIGMOD International Conference on Management of Data. New York, NY: ACM: 25–36.
- Kamel, Hosam* (2013): Visual Studio 2013 Static Code Analysis in depth: What? When and How? URL: <http://weblogs.asp.net/hosamkamel/visual-studio-2013-static-code-analysis-in-depth-what-when-and-how>, Abruf am 29.03.2015.
- Kaplan, Andreas M./Haenlein, Michael* (2010): Users of the world, unite! In: Business horizons, 53 (1): 59–68.
- Katz, Michael L./Shapiro, Carl* (1994): Systems Competition and Network Effects. In: The Journal of Economic Perspectives, 8 (2): 93–115.
- Kaufmann, Jens/Chamoni, Peter* (2013): Empowering Collaborative Business Intelligence by the use of Online Social Networks. In: Krempels, Karl-Heinz/Stocker, Alexander (Hrsg.): Proceedings of the 9th International Conference on Web Information Systems and Technologies (WEBIST). O.O.: SciTePress: 125–128.
- Kaufmann, Jens/Chamoni, Peter* (2014): Structuring Collaborative Business Intelligence: A Literature Review. In: IEEE (Hrsg.): Proceedings of the 47th Hawaii International Conference on System Sciences (HICSS '47). Los Alamitos, CA: IEEE: 3738–3747.
- Kehlenbeck, Matthias/Breitner, Michael H.* (2009): Ontology-Based Exchange and Immediate Application of Business Calculation Definitions for Online Analytical Processing. In: Pedersen, Torben B./Mohania, Mukesh K./Tjoa, A. Min (Hrsg.): Data Warehousing and Knowledge Discovery. Berlin, Heidelberg: Springer: 298–311.
- Kehua, Yang/Manirakiza, Agnes* (2012): Efficient and Semantic OLAP Aggregation Queries in a Peer to Peer Network. In: International Journal of Information and Electronics Engineering, 2 (5): 697–701.
- Keller, Andres* (2011): Breitbandkabel und Zugangsnetze. Berlin, Heidelberg: Springer.
- Kementsietsidis, Anastasios/Arenas, Marcelo/Miller, Renée J.* (2003): Mapping data in peer-to-peer systems. In: Papakonstantinou, Yannis/Halevy, Alon/Ives, Zachary (Hrsg.): Proceedings of the 2003 ACM SIGMOD international conference on Management of data. New York, NY: ACM: 325–336.
- Kemper, Hans-Georg/Baars, Henning/Mehanna, Walid* (2010): Business Intelligence - Grundlagen und praktische Anwendungen, 3. Aufl. Wiesbaden: Vieweg+Teubner.



- Kemper, Hans-Georg/Haußmann, Clemens/Lachenmaier, Jens/Lasi, Heiner* (2014): Analyse Digitaler Produktmodelle in Business-Intelligence-Infrastrukturen (DIPRO-BI). Vortragsdokumentation Expertenworkshop. Stuttgart, 22.05.2014.
- Keßler, Heinrich/Winkelhofer, Georg* (2004): Projektmanagement, 4. Aufl. Berlin, Heidelberg: Springer.
- Killich, Stephan* (2011): Formen der Unternehmenskooperation. In: Becker, Thomas/Dammer, Ingo/Howaldt, Jürgen/Loose, Achim (Hrsg.): Netzwerkmanagement. Berlin, Heidelberg: Springer: 13–22.
- Kim, Kibae/Altmann, Jörn/Hwang, Junseok* (2011): An Analysis of the Openness of the Web2.0 Service Network Based on Two Sets of Indices for Measuring the Impact of Service Ownership. In: IEEE (Hrsg.): Proceedings of the 44th Hawaii International Conference on System Sciences (HICSS '44). Los Alamitos, CA: IEEE: 1–11.
- Kimball, Ralph* (1996): The data warehouse toolkit. New York: Wiley.
- Kimball, Ralph/Ross, Margy* (2002): The data warehouse toolkit, 2. Aufl. New York, NY: John Wiley & Sons.
- Kirkman, Stephen S./Dezhgosha, Kamyar* (2011): Security review of P2P applications and networks. In: Arabnia, Hamid R./Clincy, Victor A./Solo, Ashu M. G. (Hrsg.): Proceedings of the 2011 International Conference on Internet Computing, ICOMP 2011. O.O.: CSREA Press.
- Kirsch, Werner* (1971): Entscheidungen in Organisationen. Wiesbaden: Gabler.
- Klarmann, Axel/Roth, Martin/Franczyk, Bogdan* (2013): Entscheidungsunterstützung in logistischen Netzwerken - Ansätze für eine Umsetzung. In: Baars, Henning (Hrsg.): WSBI-2012. Stuttgart: CEUR-WS: 28–36.
- Klein, Dominik/Tran-Gia, Phuoc/Hartmann, Matthias* (2013): Big Data. In: Informatik Spektrum, 36 (3): 319–323.
- Klement, Erich P./Mesiar, Radko/Pap, Endre* (2005): Triangular norms: Basic notions and properties. In: Klement, Erich P./Mesiar, Radko (Hrsg.): Logical, algebraic, analytic, and probabilistic aspects of triangular norms. Amsterdam: Elsevier: 17–60.
- Knight, Shirlee-Ann/Burn, Janice* (2005): Developing a Framework for Assessing Information Quality on the World Wide Web. In: Informing Science Journal, 8 (5): 159–172.
- Knoke, David/Yang, Song* (2008): Social network analysis, 2. Aufl. Los Angeles et al.: Sage.
- Knoppe, Marc* (1997): Strategische Allianzen in der Kreditwirtschaft. München, Wien: Oldenbourg.
- Koch, Susanne* (2011): Einführung in das Management von Geschäftsprozessen. Berlin, Heidelberg: Springer.
- Kohn, Wolfgang* (2005): Statistik. Berlin, Heidelberg: Springer.
- Kommission der Europäischen Gemeinschaften* (2003): Empfehlung der Kommission vom 6. Mai 2003 betreffend die Definition der Kleinstunternehmen sowie der kleinen und mittleren Unternehmen. In: Amtsblatt der Europäischen Union Nr. L 124/36: 1–6.

- Kornmeier, Martin* (2007): Wissenschaftstheorie und wissenschaftliches Arbeiten. Heidelberg: Physica.
- Krcmar, Helmut* (2010): Informationsmanagement, 5. Aufl. Berlin, Heidelberg: Springer.
- Krivitsky, Pavel N./Handcock, Mark S./Raftery, Adrian E./Hoff, Peter D.* (2009): Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models. In: Social Networks, 31 (3): 204–213.
- Krneta, Dragoljub/Jovanovic, Vladan/Marjanovic, Zoran* (2014): A direct approach to physical Data Vault design. In: Computer Science and Information Systems, 11 (2): 569–599.
- Kuhn, Axel/Hellingrath, Bernd* (2002): Supply-Chain-Management. Berlin, Heidelberg: Springer.
- Kurbel, Karl E.* (2013): Enterprise Resource Planning and Supply Chain Management. Dordrecht: Springer.
- LAB & Company* (2012): 26. LAB Managerpanel. URL: <http://www.labcompany.net/de/press/releases/2012/144/>, Abruf am 15.11.2012.
- Ladley, John* (2010): Making EIM Enterprise Information Management Work for Business. Boston: Morgan Kaufmann.
- Lange, Kersten* (2010): Kooperationen in der Automobilindustrie – Analyse und Systematisierung. Westfälische Wilhelms-Universität Münster, Münster.
- Lange, Kersten* (2011): Kooperationen in der deutschen Automobilindustrie: Ergebnisse einer empirischen Analyse. Westfälische Wilhelms-Universität Münster, Münster.
- Laudon, Kenneth C./Laudon, Jane P.* (2006): Management information systems, 9. Aufl. Upper Saddle River, NJ: Pearson Prentice Hall.
- Law, Averill M.* (2007): Simulation modeling and analysis, 4. Aufl. Boston et al.: McGraw-Hill.
- Leimeister, Jan M.* (2014): Collaboration Engineering. Berlin, Heidelberg: Springer.
- Levera, Jorge/Barán, Benjamin/Grossman, Robert* (2004): Experimental Studies Using Median Polish Procedure to Reduce Alarm Rates in Data Cubes of Intrusion Data. In: Chen, Hsinchun/Moore, Reagan/Zeng, Daniel D./Leavitt, John (Hrsg.): Intelligence and Security Informatics. Berlin, Heidelberg: Springer: 457–466.
- Liebscher, Anna K.* (2012): Betriebliche Ressourcensicherung durch Nachhaltigkeitskooperationen. Berlin, Münster: LIT.
- Liggesmeyer, Peter* (2009): Software-Qualität, 2. Aufl. Heidelberg: Spektrum.
- Lindemann, Felix* (2005): Quo vadis Supply Chain Management. In: Aliche, Knut (Hrsg.): Planung und Betrieb von Logistiknetzwerken. 2. Aufl. Berlin, Heidelberg: Springer: 281–326.
- Linstedt, Dan E.* (2002): Data Vault Series 1 - Data Vault Overview. In: Seiner, Robert S. (Hrsg.): The Data Administration Newsletter. URL: <http://www.tdan.com/view-articles/5054/>, Abruf am 09.12.2014.

- Lipinski, Klaus* (2015): CBI (collaborative business intelligence). URL: <http://www.computer.org/csdl/trans/tk/preprint/06834809-abs.html>, Abruf am 17.03.2015.
- Liu, Linghze/Daniels, Hennie* (2012): Towards a Value Model for Collaborative, Business Intelligence-supported Risk Assessment. In: Proceedings of the 6th International Workshop on Value Modeling and Business Ontology (VMBO 2012): 1–5. URL: [http://vmbo2012.isis.tuwien.ac.at/?page\\_id=4](http://vmbo2012.isis.tuwien.ac.at/?page_id=4), Abruf am 30.11.2014.
- Liu, Lingxia/Kim, Kisu/Sun, Wei* (2012): An Empirical Research of Factors Affecting Collaborative Decision Making System. In: International Journal of Hybrid Information Technology, 5 (4): 1–10.
- Lodi, Stefano/Penzo, Wilma/Mandreoli, Federica/Martoglia, Riccardo/Sassatelli, Simona* (2008): Semantic peer, here are the neighbors you want! In: Kemper, Alfons (Hrsg.): Proceedings of the 11th international conference on Extending database technology (EDBT '08). New York, NY: ACM: 26–37.
- Loos, Peter/Mettler, Tobias/Winter, Robert/Goeken, Matthias/Frank, Ulrich/Winter, Alfred* (2013): Methodenpluralismus in der Wirtschaftsinformatik? In: WIRTSCHAFTSINFORMATIK, 55 (6): 457–464.
- Ludewig, Jochen/Lichter, Horst* (2010): Software-Engineering, 2. Aufl. Heidelberg: Dpunkt.
- Luhn, Hans P.* (1958): A Business Intelligence System. In: IBM Journal of Research and Development, 2 (4): 314–319.
- Magin, Vera/Heil, Oliver P./Fürst, Ronny A.* (2005): Kooperation und Competition: Erklärungsperspektive der Spieltheorie. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. 2. Aufl. Wiesbaden: Gabler: 121–140.
- Mahboubi, Hadj/Darmont, Jérôme* (2009): Enhancing XML data warehouse query performance by fragmentation. In: Shin, Sung Y./Ossowski, Sascha (Hrsg.): Proceedings of the 2009 ACM symposium on Applied Computing (SAC '09). New York, NY: ACM: 1555–1562.
- Maier, Ronald/Hädrich, Thomas* (2006): Centralized versus peer-to-peer knowledge management systems. In: Knowledge and Process Management, 13 (1): 47–61.
- Makowski, Louis/Ostroy, Joseph M.* (2001): Perfect Competition and the Creativity of the Market. In: Journal of Economic Literature, 39 (2): 479–535.
- Malaverri, Joana E. G./Mota, Matheus S./Medeiros, Claudia M. B.* (2013): Estimating the quality of data using provenance: a case study in eScience. In: Association for Information Systems (Hrsg.): Proceedings of the 19th Americas Conference on Information Systems 2013 (AMCIS). Red Hook, NY: Curran: 1–10.
- Malek, Mirosław* (2006): A Perspective on Parallel and Distributed Computing. In: Reisig, Wolfgang/Freytag, Johann C. (Hrsg.): Informatik. Berlin, Heidelberg: Springer: 197–220.
- Mandreoli, Federica/Martoglia, Riccardo/Sassatelli, Simona/Penzo, Wilma* (2006): SRI: exploiting semantic information for effective query routing in a PDMS. In: Bonifati, Angela/Fundulaki, Irini (Hrsg.): Proceedings of the



- 8th ACM International Workshop on Web Information and Data Management (WIDM 2006). New York, NY: ACM: 19–26.
- Manshaei, Mohammad H./Zhu, Quanyan/Alpcan, Tansu/Bacşar, Tamer/Hubaux, Jean-Pierre* (2013): Game theory meets network security and privacy. In: *ACM Computing Surveys*, 45 (3): 1–39.
- Manzano, Marc/Calle, Eusebi/Harle, David* (2011): Quantitative and qualitative network robustness analysis under different multiple failure scenarios. In: *IEEE* (Hrsg.): 3rd International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT). Budapest, Hungary: IEEE.
- Manzano, Marc/Calle, Eusebi/Torres-Padrosa, Victor/Segovia, Juan/Harle, David* (2013): Endurance: A new robustness measure for complex networks under multiple failure scenarios. In: *Computer Networks*, 57 (17): 3641–3653.
- Markus, M. L./Silver, Mark S.* (2008): A Foundation for the Study of IT Effects: A New Look at DeSanctis and Poole’s Concepts of Structural Features and Spirit. In: *Journal of the Association for Information Systems*, 9 (10): 609–633.
- Martins, Valério A.* (2013): Arquitetura de um ambiente colaborativo de business intelligence baseado em um repositório de ontologias e serviços de dados. Brasília: Universidade de Brasília.
- Martins, Valério A./daCosta, Joao P. L./deSousa, Rafael T.* (2012): Architecture of a Collaborative Business Intelligence Environment based on an Ontology Repository and Distributed Data Services. In: *INSTICC* (Hrsg.): Proceedings of the 4th International Conference on Knowledge Management and Information Sharing (KMIS 2012). Barcelona: INSTICC: 99–106.
- Marwell, Gerald/Oliver, Pamela E./Prahl, Ralph* (1988): Social Networks and Collective Action: A Theory of the Critical Mass. III. In: *American Journal of Sociology*, 94 (3): 502–534.
- Matt, Christian* (2012): Groupware. In: *Zeitschrift für Controlling und Management*, 56 (3): 164–167.
- Maur, Eitel von* (2009): Konstruktivismus und Wirtschaftsinformatik - Begriffsver(w)irrungen. In: *Becker, Jörg/Krcmar, Helmut/Niehaves, Björn* (Hrsg.): *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*. Heidelberg: Physica: 133–160.
- Mawlood-Yunis, Abdul-Rahman/Weiss, Michael/Santoro, Nicola* (2011): A Reference Model for Semantic Peer-to-Peer Networks. In: *Spaccapietra, Stefano* (Hrsg.): *Journal on Data Semantics XV*. Berlin, Heidelberg: Springer: 37–65.
- Mazón, Jose-Norberto/Lechtenböcker, Jens/Trujillo, Juan* (2009): A survey on summarizability issues in multidimensional modeling. In: *Data & Knowledge Engineering*, 68 (12): 1452–1469.
- McAfee, Andrew* (2006): Enterprise 2.0, version 2.0. URL: [http://andrewmcafee.org/2006/05/enterprise\\_20\\_version\\_20/](http://andrewmcafee.org/2006/05/enterprise_20_version_20/), Abruf am 06.12.2013.
- McBrien, Peter/Poulovassilis, Alexandra* (2003): Data Integration by Bi-Directional Schema Transformation Rules. In: *Dayal,*

- Umeshwar/Ramamritham, Krithi/Vijayaraman, Thodukattil M. (Hrsg.): Proceedings of the 19th International Conference on Data Engineering (ICDE'03). Los Alamitos, CA: IEEE: 227–238.
- McBrien, Peter/Poulovassilis, Alexandra* (2004): Defining Peer-to-Peer Data Integration Using Both as View Rules. In: Aberer, Karl/Koubarakis, Manolis/Kalogeraki, Vana (Hrsg.): Databases, Information Systems, and Peer-to-Peer Computing. Berlin, Heidelberg: Springer: 91–107.
- McDonald, Kevin* (2002): Mastering the SAP business information warehouse. New York: John Wiley & Sons.
- McKnight, William* (2014): Information Management. Boston: Morgan Kaufmann.
- McPherson, Miller/Smith-Lovin, Lynn/Cook, James M.* (2001): Birds of a Feather: Homophily in Social Networks. In: Annual Review of Sociology, 27: 415–444.
- Medhi, Deepankar/Ramasamy, Karthikeyan* (2007): Network routing. Amsterdam et al.: Elsevier; Morgan Kaufmann.
- Mehrwald, Christian* (2010): Datawarehousing mit SAP BW 7, 5. Aufl. Heidelberg: Dpunkt.
- Meier, John J./Conkling, Thomas W.* (2008): Google Scholar's Coverage of the Engineering Literature: An empirical study. In: The Journal of Academic Librarianship, 34 (3): 196–201.
- Meinel, Christoph/Sack, Harald* (2012): Internetworking. Berlin, Heidelberg: Springer.
- Melkas, Helinä* (2004): Analyzing Information Quality In Virtual Service Networks With Qualitative Interview Data. In: M.I.T. Information Quality Program (Hrsg.): Proceedings of the 9th International Conference on Information Quality (ICIQ-04). Cambridge, MA: Massachusetts Institute of Technology: 74–88.
- Mell, Peter/Grance, Tim* (2011): The NIST Definition of Cloud Computing. Technical Report, National Institute of Standards and Technology. URL: <http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf>, Abruf am 30.11.2014.
- Melnik, Sergey/Garcia-Molina, Hector/Rahm, Erhard* (2002): Similarity flooding: a versatile graph matching algorithm and its application to schema matching. In: Georgakopoulos, Dimitrios/Agrawal, Rakesh/Dittrich, Klaus (Hrsg.): Proceedings of the 18th International Conference on Data Engineering (ICDE'02). Los Alamitos, CA: IEEE: 117–128.
- Mena, Eduardo/Kashyap, Vipul/Illarramendi, Arantza/Sheth, Amith P.* (2000): Imprecise Answers in Distributed Environments. In: International Journal of Cooperative Information Systems, 9 (4): 403–425.
- Mertens, Peter* (1982): Simulation, 2. Aufl. Stuttgart: Poeschel.
- Mertens, Peter* (1995): Wirtschaftsinformatik: Von den Moden zum Trend. In: König, Wolfgang (Hrsg.): Wirtschaftsinformatik '95. Heidelberg: Physica: 25–64.

- Mertens, Peter* (2006): Moden und Nachhaltigkeit in der Wirtschaftsinformatik. Arbeitsbericht des Lehrstuhls für BWL, insbes. Wirtschaftsinformatik I. Universität Erlangen-Nürnberg, Erlangen-Nürnberg.
- Mertens, Peter/Bodendorf, Freimut/König, Wolfgang/Picot, Arnold/Schumann, Matthias/Hess, Thomas* (2012): Grundzüge der Wirtschaftsinformatik. Berlin, Heidelberg: Springer.
- Mertens, Peter/Meier, Marco C.* (2009): Integrierte Informationsverarbeitung 2, 10. Aufl. Wiesbaden: Gabler.
- Mettler, Tobias/Pinto, Roberto/Raber, David* (2012): An Intelligent Supply Chain Design for Improving Delivery Reliability. In: International Journal of Information Systems and Supply Chain Management, 5 (2): 1–20.
- Mettler, Tobias/Raber, David* (2011): Developing a collaborative business intelligence system for improving delivery reliability in business networks. In: Thoben, Klaus-Dieter/Stich, Volker/Imtiaz, Ali (Hrsg.): Proceedings of the 17th International Conference on Concurrent Enterprising (ICE'11). Aachen: FIR e. V. at RWTH Aachen University: 1–7.
- Michel, Luis M.* (2009): Management von Kooperationen im Bereich Forschung und Entwicklung. Eine empirische Studie. In: Manz, Carsten (Hrsg.): Konstanzer Managementschriften. Konstanz: Hochschule Konstanz.
- Microsoft* (2013): Visual Studio 2013: Was ist statische Code-Analyse? URL: [http://www.microsoft.com/germany/visualstudio/aktuell/news/show.aspx?id=msdn\\_de\\_51099](http://www.microsoft.com/germany/visualstudio/aktuell/news/show.aspx?id=msdn_de_51099), Abruf am 29.03.2015.
- Miller, Michael* (2001): Discovering P2P. San Francisco: Sybex.
- Minuto Espil, Mauricio/Vaisman, Alejandro A.* (2004): Aggregate queries in peer-to-peer OLAP. In: Song, Il-Yeol/Davis, Karen C. (Hrsg.): Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP (DOLAP). New York, NY: ACM: 102–111.
- Mislove, Alan/Roughan, Matthew/Marcon, Massimiliano/Gummadi, Krishna P./Druschel, Peter/Bhattacharjee, Bobby* (2007): Measurement and analysis of online social networks. In: ACM (Hrsg.): Proceedings of the 2007 ACM SIGCOMM Internet Measurement Conference (ICM '07). New York, NY: ACM: 29–42.
- Molensky, Lydia/Ketter, Wolfgang/Collins, John/Bloemhof-Ruwaard, Jacqueline M./Koppel, Hans van de* (2010): Business intelligence gap analysis: A user, supplier and academic perspective. In: Bui, Tung/Jarke, Matthias (Hrsg.): Proceedings of the 12th International Conference on Electronic Commerce (ICEC '10). New York, NY: ACM: 119–128.
- Morin, Nancy R.* (1997): The flora of North America project. In: The American Biology Teacher, 59 (6): 338–343.
- Morschett, Dirk* (2005): Formen von Kooperationen, Allianzen und Netzwerken. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. 2. Aufl. Wiesbaden: Gabler: 377–404.
- Mucksch, Harry/Behme, Wolfgang* (2000): Das Data Warehouse-Konzept als Basis einer unternehmensweiten Informationslogistik. In: Mucksch, Harry/Behme, Wolfgang (Hrsg.): Das Data Warehouse-Konzept. 4. Aufl. Wiesbaden: Gabler: 3–80.

- Müller, Dietmar* (2013): Die drei Analytics-Toprends: "Big Data", "Self-Service" und "Mobile". URL: <http://ibmexperts.computerwoche.de/analytics-big-data/artikel/die-drei-analytics-toptrends-big-data-self-service-und-mobile>, Abruf am 21.10.2014.
- Muntean, Mihaela* (2012a): Business Intelligence Approaches. In: *Mathematical Models & Methods in Applied Sciences*, 1: 192–196.
- Muntean, Mihaela* (2012b): Theory and Practice in Business Intelligence. URL: <http://ssrn.com/abstract=2144440>, Abruf am 21.04.2015.
- Nagl, Wolf* (1999): Wirtschaftsinformatik: Rück- und Ausblick eines Praktikers. In: *Becker, Jörg* (Hrsg.): *Wirtschaftsinformatik und Wissenschaftstheorie*. Wiesbaden: Gabler: 417–448.
- Nakagawa, Elisa Yumi/Oliveira Antonino, Pablo/Becker, Martin* (2011): Reference Architecture and Product Line Architecture: A Subtle But Critical Difference. In: *Crnkovic, Ivica/Gruhn, Volker/Book, Matthias* (Hrsg.): *Software Architecture*. Berlin, Heidelberg: Springer: 207–211.
- Naumann, Felix/Leser, Ulf/Freytag, Johann C.* (1999): Quality-driven Integration of Heterogenous Information Systems. In: *Atkinson, Malcolm P./Orlowska, Maria E./Valduriez, Patrick/Zdonik, Stanley B./Brodie, Michael L.* (Hrsg.): *Proceedings of the 25th International Conference on Very Large Data Bases (VLDB '99)*. San Francisco, CA: Morgan Kaufmann: 447–458.
- Navrade, Frank* (2008): *Strategische Planung mit Data-Warehouse-Systemen*. Wiesbaden: Gabler.
- Neapolitan, Richard E.* (2014): *Foundations of algorithms*, 5. Aufl. Burlington, MA: Jones & Bartlett.
- Nelson, Gregory S.* (2010): Business Intelligence 2.0: Are we there yet. In: *SAS* (Hrsg.): *Proceedings of the SAS Global Forum*. URL: <http://support.sas.com/resources/papers/proceedings10/TOC.html>, Abruf am 30.11.2014.
- Neuhäuser, Markus* (2014): Wilcoxon–Mann–Whitney Test. In: *Lovric, Miodrag* (Hrsg.): *International Encyclopedia of Statistical Science*. Berlin, Heidelberg: Springer: 1656–1658.
- Neumann, John von/Morgenstern, Oskar* (2004): *Theory of games and economic behavior*, 60. Aufl. Princeton, NJ: Princeton Univ. Press.
- Neumayr, Bernd/Schrefl, Michael/Thalheim, Bernhard* (2010): Heterogeneous hierarchies in data warehouses. In: *Link, Sebastian/Ghose, Aditya* (Hrsg.): *Proceedings of the 6th Asia-Pacific Conference on Conceptual Modelling*. Darlington: Australian Computer Society: 61–70.
- Newman, Mark E. J.* (2003): The Structure and Function of Complex Networks. In: *SIAM Review*, 45 (2): 167–256.
- Nguyen, Hieu V./Bai, Li* (2011): Cosine Similarity Metric Learning for Face Verification. In: *Kimmel, Ron/Klette, Reinhard/Sugimoto, Akihiro* (Hrsg.): *Computer Vision – ACCV 2010*. Berlin, Heidelberg: Springer: 709–720.
- Nobari, Sadegh/Lu, Xuesong/Karras, Panagiotis/Bressan, Stéphane*: Fast random graph generation. In: *Ailamaki, Anastasia/Amer-Yahia, Sihem/Pate, Jignesh/Risch, Tore/Senellart, Pierre/Stoyanovich, Julia* (Hrsg.): *Proceed-*

- ings of the 14th International Conference on Extending Database Technology (EDBT/ICDT '11). New York, NY, USA: ACM: 331–342.
- Nunamaker, Jay F./Dennis, Alan R./Valacich, Joseph S./Vogel, Douglas R./George, Joey R.* (1991): Electronic Meeting Systems to Support Group Work. In: *Communications of the ACM*, 34 (7): 40–61.
- Object Management Group* (2010): OMG Unified Modeling Language (OMG UML), Infrastructure Version 2.4.1. URL: <http://www.omg.org/spec/UML/2.4.1/Infrastructure/PDF/>, Abruf am 27.03.2015.
- Ochs, Carsten/Löw, Martina* (2012): Un/Faire Informationspraktiken: Internet Privacy aus Sozialwissenschaftlicher Perspektive. In: *Buchmann, Johannes* (Hrsg.): *Internet Privacy*. Berlin, Heidelberg: Springer: 15–62.
- Olaru, Marius-Octavian* (2012): Partial Multi-dimensional Schema Merging in Heterogeneous Data Warehouses. In: *Atzeni, Paolo/Cheung, David/Sudha, Ram* (Hrsg.): *Conceptual modeling: 31st international conference, ER 2012*. Berlin, Heidelberg: Springer: 563–571.
- Ooi, BengChin/Liau, ChuYee/Tan, Kian-Lee* (2003): Managing Trust in Peer-to-Peer Systems Using Reputation-Based Techniques. In: *Dong, Guozhu/Tang, Changjie/Wang, Wei* (Hrsg.): *Advances in Web-Age Information Management*. Berlin, Heidelberg: Springer: 2–12.
- OR/MS Today* (2013): Simulation Software Survey. URL: <http://www.orms-today.org/surveys/Simulation/Simulation1.html>, Abruf am 28.03.2015.
- O'Reilly, Tim* (2005): What is Web 2.0? URL: <http://oreilly.com/web2/archive/what-is-web-20.html>, Abruf am 11.05.2015.
- Oren, Shmuel S./Smith, Stephen A.* (1981): Critical Mass and Tariff Structure in Electronic Communications Markets. In: *The Bell Journal of Economics*, 12 (2): 467–487.
- Osiecka, Agnieszka* (2006): *Grenzüberschreitende Unternehmenskooperationen*. Wiesbaden: DUV.
- Österle, Hubert/Becker, Jörg/Frank, Ulrich/Hess, Thomas/Karagiannis, Dimitris/Krcmar, Helmut/Loos, Peter/Merstens, Peter/Oberweis, Andreas/Sinz, Elmar J.* (2010): Memorandum zur gestaltungsorientierten Wirtschaftsinformatik. In: *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 62 (6): 664–672.
- Oxford Dictionaries* (2015): Definition of collaboration in English. URL: <http://www.oxforddictionaries.com/definition/english/collaboration>, Abruf am 11.05.2015.
- Panja, Rajarsi/Paul, Subrata* (2014): A Review Report on the Evolution and Implementation of Business Intelligence Technique in the Banking Sector. In: *International Journal of Computer Science and Mobile Applications (IJCSMA)*, 2 (11): 109–114.
- Paprottka, Stephan* (1996): *Unternehmenszusammenschlüsse*. Wiesbaden: Gabler.
- Pardillo, Jesus/Mazon, Jose N.* (2011): Using Ontologies for the Design of Data Warehouses. In: *International Journal of Database Management Systems*, 3 (2): 73–87.



- Park, Nam H./Joo, Kil H.* (2013): Query Processing and Index Structure on Cloud Peer-to-Peer OLAP. In: Hassanien, Aboul E./Choi, Jin-Tak (Hrsg.): Current Research on HCI and Multimedia. Advanced Science and Technology Letters (ASTL 43). Korea: Catholic University of DaeGu: 43–46.
- Paul, Joachim* (2014): Trends und Megatrends in Business Intelligence. In: Controlling & Management Review, 58 (1): 69.
- Pedersen, Torben B.* (2013): Managing Complex Multidimensional Data. In: Auffaure, Marie-Aude/Zimányi, Esteban (Hrsg.): Business Intelligence. Berlin, Heidelberg: Springer: 1–28.
- Peffers, Ken/Tuunanen, Tuure/Rothenberger, Marcus A./Chatterjee, Samir* (2007): A Design Science Research Methodology for Information Systems Research. In: Journal of Management Information Systems, 24 (3): 45–77.
- PewResearch* (2013): Social Networking Fact Sheet. URL: <http://www.pewinternet.org/fact-sheets/social-networking-fact-sheet/>, Abruf am 12.03.2014.
- Pfähler, Wilhelm/Wiese, Harald* (2008): Unternehmensstrategien im Wettbewerb. Berlin, Heidelberg: Springer.
- Pfeiffer, Daniel/Niehaves, Björn* (2005): Evaluation of Conceptual Models - A Structuralist Approach. In: Bartmann, Dieter/Rajola, Federico/Kallinikos, Jannis/Avison, David E./Winter, Robert/Ein-Dor, Phillip/Becker, Jörg/Bodendorf, Freimut/Weinhardt, Christof (Hrsg.): Proceedings of the 13th European Conference on Information Systems (ECIS 2005). Regensburg: Universität Regensburg: 459–470.
- Pfohl, Hans-Christian/Braun, Günther E.* (1981): Entscheidungstheorie. Landsberg am Lech: Moderne Industrie.
- Picot, Arnold* (1991): Ökonomische Theorien der Organisation. In: Ordelheide, Dieter/Rudolph, Bernd/Büsselmann, Elke (Hrsg.): Betriebswirtschaftslehre und ökonomische Theorie. Stuttgart: Poeschel: 143–170.
- Poole, Marshall S./Holmes, Michael./Watson, Richard/DeSanctis, Gerardine* (1993): Group Decision Support Systems and Group Communication. In: Communication Research, 20 (2): 176–213.
- Porter, Michael E.* (2014): Wettbewerbsvorteile, 8. Aufl. Frankfurt am Main: Campus.
- Power, Daniel J.* (2009): A Brief History of Decision Support Systems, version 4.1. URL: <http://dssresources.com/history/dsshhistory.html>, Abruf am 17.11.2013.
- Prat, Nicolas/Madnick, Stuart E.*: Evaluating and Aggregating Data Believability across Quality Sub-Dimensions and Data Lineage. MIT Sloan School Working PaPer 4670-07. In: Kothari, S. P. (Hrsg.): MIT Sloan School Working Papers. Cambridge, MA: MIT Sloan School of Management: 1–6.
- Priebe, Torsten/Pernul, Günther* (2000): Towards OLAP security design - survey and research issues. In: Missaoui, Rokia/Song, Il-Yeol (Hrsg.): Proceedings of the 3rd ACM international workshop on Data warehousing and OLAP (DOLAP '00). New York, NY: ACM: 33–40.

- Provan, Keith G./Fish, Amy/Sydow, Jörg* (2007): Interorganizational Networks at the Network Level: A Review of the Empirical Literature on Whole Networks. In: *Journal of Management*, 33: 479–516.
- Rabe, Markus* (1998): Einführung. In: Kuhn, Axel/Rabe, Markus (Hrsg.): *Simulation in Produktion und Logistik*. Berlin et al.: Springer: 1–10.
- Rabiner, Lawrence R.* (1988): A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. In: *IEEE* (Hrsg.): *Proceedings of IEEE*. Los Alamitos, CA: IEEE: 257–286.
- Rahm, Erhard/Bernstein, Philip A.* (2001): A Survey of Approaches to Automatic Schema Matching. In: *The VLDB Journal*, 10 (4): 334–350.
- Ramanathan, Kannan/Seth, Anju/Thomas, Howard* (1997): Explaining Joint Ventures. In: Beamish, Paul W./Killing, J. P. (Hrsg.): *Cooperative strategies*. San Francisco: New Lexington Press: 51–85.
- Riazati, Dariush/Thom, James A./Zhang, Xiuzhen* (2011): Enforcing strictness in integration of dimensions: beyond instance matching. In: Song, Il-Yeol/Cuzzocrea, Alfredo/Davis, Karen C. (Hrsg.): *Proceedings of the 14th ACM International Workshop on Data Warehousing and OLAP (DOLAP)*. New York, NY: ACM: 9–16.
- Richter, Daniel/Riemer, Kai/Brocke, Jan vom* (2011): Internet Social Networking. In: *WIRTSCHAFTSINFORMATIK*, 53 (2): 89–103.
- Riege, Christian/Saat, Jan/Bucher, Tobias* (2009): Systematisierung von Evaluationsmethoden in der gestaltungsorientierten Wirtschaftsinformatik. In: Becker, Jörg/Krcmar, Helmut/Niehaves, Björn (Hrsg.): *Wissenschaftstheorie und gestaltungsorientierte Wirtschaftsinformatik*. Heidelberg: Physica: 69–86.
- Rijsbergen, Cornelis J. van* (1979): *Information retrieval*, 2. Aufl. London, Boston: Butterworths.
- Rizzi, Stefano* (2012): Collaborative Business Intelligence. In: Aufaure, Marie-Aude/Zimányi, Esteban (Hrsg.): *Business Intelligence*. Berlin, Heidelberg: Springer: 186–205.
- Roe, Charles* (2011): Business Intelligence 3.0 – Social Analytics Part 1. URL: <http://www.dataversity.net/business-intelligence-3-0-social-analytics-part-1/6309/>, Abruf am 27.09.2012.
- Rohrmoser, Klaus* (2014): Optimierung von Entscheidungsprozessen mit Collaborative BI: Steter Austausch von Wissen und Erfahrungen. In: *BI Spektrum*, 9 (1): 28–30.
- Rosemann, Michael/Eggert, Mathias/Voigt, Matthias/Beverungen, Daniel* (2012): Leveraging Social Network Data for Analytical CRM Strategies - The Introduction of Social BI. Paper 95. In: Wareham, Jonathan/Busquets, Xavier/Valor, Josep/Sieber, Sandra (Hrsg.): *Proceedings of the 20th European Conference on Information Systems (ECIS)*. Barcelona: Escuela Superior de Administración y Dirección de Empresas: 1–12.
- Rosemann, Michael/Schütte, Reinhard* (1997): Grundsätze ordnungsmäßiger Referenzmodellierung. In: Becker, Jörg/Grob, Heinz L./Müller-Funk, Ulrich/Vossen, Georg (Hrsg.): *Entwicklungsstand und Entwicklungsperspektiven der Referenzmodellierung*. Münster: Institut für Wirtschaftsinformatik, WWU Münster: 16–33.



- Roughgarden, Tim/Tardos, Éva* (2002): How bad is selfish routing? In: Journal of the ACM, 49 (2): 236–259.
- Rouse, Margaret* (2012): collaborative BI (collaborative business intelligence). URL: <http://searchbusinessanalytics.techtarget.com/definition/collaborative-BI-collaborative-business-intelligence>, Abruf am 11.12.2013.
- Roy, Geoffrey G.* (2006): Designing and explaining programs with a literate pseudocode. In: Journal on Educational Resources in Computing, 6 (1): 1–es.
- Rubin, Jeffrey Z.* (1984): Introduction. In: Swap, Walter C. (Hrsg.): Group decision making. Beverley Hills, CA: Sage: 15–44.
- Rupprecht-Däullary, Marita* (1994): Zwischenbetriebliche Kooperation. Wiesbaden: DUV.
- Saito, Takamichi/Sekiguchi, Kiyomi/Hatsugai, Ryosuke* (2008): Authentication Binding between TLS and HTTP. In: Takizawa, Makoto/Barolli, Leonard/Enokido, Tomoya (Hrsg.): Network-Based Information Systems. Berlin, Heidelberg: Springer: 252–262.
- Sapia, Carsten* (2000): PROMISE: Predicting Query Behavior to Enable Predictive Caching Strategies for OLAP Systems. In: Kambayashi, Yohiko/Mohania, Mukesh/Tjoa, A. Min (Hrsg.): Proceedings of the 2nd International Conference on Data Warehousing and Knowledge Discovery (DAWAK). London: Springer: 224–233.
- Scannapieco, Monica/Missier, Paolo/Batini, Carlo* (2005): Data Quality at a Glance. In: Datenbank-Spektrum, 14: 6–14.
- Schauer, Carola* (2011): Die Wirtschaftsinformatik im internationalen Wettbewerb. Wiesbaden: Gabler.
- Schief, Markus/Buxmann, Peter/Schiereck, Dirk* (2013): Fusionen und Unternehmensübernahmen in der Softwareindustrie. In: WIRTSCHAFTSINFORMATIK, 55 (6): 421–433.
- Schlagwein, Daniel/Schoder, Detlef/Fischbach, Kai* (2010): Openness of Information Resources – A Framework-based Comparison of Mobile Platforms. Paper 163. In: Alexander, Trish/Turpin, Marita/Deventer, Juan P. van (Hrsg.): Proceedings of the 18th European Conference on Information Systems (ECIS'10). Pretoria: University of Pretoria.
- Schlegel, Kurt* (2013): Hype Cycle for Business Intelligence and Analytics, 2013, G00252553. URL: <https://www.gartner.com/doc/2570415/hype-cycle-business-intelligence-analytics>, Abruf am 30.11.2014.
- Schlegel, Kurt* (2014): Hype Cycle for Business Intelligence and Analytics, 2014, G00262070. URL: <https://www.gartner.com/doc/2813223/hype-cycle-business-intelligence-analytics>, Abruf am 05.03.2015.
- Schmidt, Axel/Kiefer, Clemens* (2005): Kooperationen zwischen mittelständischen Unternehmen. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. 2. Aufl. Wiesbaden: Gabler: 1357–1381.
- Schoder, Detlef/Fischbach, Kai* (2002): Peer-to-Peer. In: WIRTSCHAFTSINFORMATIK, 44 (6): 587–589.

- Scholta, Claudia* (2005): Erfolgsfaktoren unternehmensübergreifender Kooperation am Beispiel der mittelständischen Automobilzulieferindustrie in Sachsen. Chemnitz: Technische Universität Chemnitz.
- Schönherr, Marten* (2004): Enterprise Architecture Frameworks. In: Aier, Stephan/Schönherr, Marten (Hrsg.): Enterprise Application Integration – Serviceorientierung und nachhaltige Architekturen. Berlin: Gito: 3–48.
- Schurz, Gerhard* (2014): Ceteris Paribus and Ceteris Rectis Laws: Content and Causal Role. In: Erkenntnis, 79 (10): 1801–1817.
- Schütz, Christoph/Schrefl, Michael/Neumayr, Bernd/Sierninger, Daniel* (2011): Incremental integration of data warehouses: the hetero-homogeneous approach. In: Song, Il-Yeol/Cuzzocrea, Alfredo/Davis, Karen C. (Hrsg.): Proceedings of the 14th ACM International Workshop on Data Warehousing and OLAP (DOLAP). New York, NY: ACM: 25–30.
- Schwalm, Stephan/Bange, Carsten* (2004): Einsatzpotenziale von XML in Business-Intelligence-Systemen. In: WIRTSCHAFTSINFORMATIK, 46 (1): 5–14.
- Schwerk, Anja* (2000): Dynamik von Unternehmenskooperationen. Berlin: Duncker & Humblot.
- Seitz, Sebastian* (2012): Social Business Intelligence. In: BI Spektrum, 7 (3): 28–30.
- Sell, Axel* (2002): Internationale Unternehmenskooperationen, 2. Aufl. München, Wien: Oldenbourg.
- Seo, Hyunjeong/Kim, Wanyeon/Lee, Jihyung/Youn, Buhyun* (2013): Network-based approaches for anticancer therapy (Review). In: International journal of oncology, 43 (6): 1737–1744.
- Sepucha, Karen R./Belkora, Jeffrey K./Tripathy, Debasish/Esserman, Laura J.* (2000): Building Bridges Between Physicians and Patients. In: Journal of Clinical Oncology, 18 (6): 1230–1238.
- Seshadri, Sangeetha/Cooper, Brian F./Liu, Ling* (2005): CubeCache: Efficient and Scalable Processing of OLAP Aggregation Queries in a Peer-to-Peer Network. Technical Report GIT-CERCS-07-12. URL: <http://hdl.handle.net/1853/20072>, Abruf am 03.11.2014.
- Shah, Neepa K.* (2014): Data warehouse systems in the environment of Cloud Computing. In: IEEE (Hrsg.): Proceedings of the 2014 International Conference on Computing for Sustainable Global Development (INDIACom). Los Alamitos, CA: IEEE: 118–123.
- Shen, Ying* (2014): Global Internet Routing Table Reaches 512k Milestone. URL: <http://blogs.cisco.com/sp/global-internet-routing-table-reaches-512k-milestone>, Abruf am 25.02.2015.
- Shvaiko, Pavel/Euzenat, Jérôme* (2005): A Survey of Schema-Based Matching Approaches. In: Spaccapietra, Stefano (Hrsg.): Journal on Data Semantics IV. Berlin, Heidelberg: Springer: 146–171.
- Shvaiko, Pavel/Euzenat, Jérôme* (2013): Ontology Matching: State of the Art and Future Challenges. In: IEEE Transactions on Knowledge and Data Engineering, 25 (1): 158–176.

- Siebertz, Karl/van Bebber, David/Hochkirchen, Thomas*: Statistische Versuchssplanung. Berlin, Heidelberg: Springer.
- Simon, Alan R./Shaffer, Steven L.* (2002): Chapter 7 - Core Technologies and Building Blocks. In: Simon, Alan R./Shaffer, Steven L. (Hrsg.): Data Warehousing And Business Intelligence For e-Commerce. San Francisco: Morgan Kaufmann: 125–171.
- Singhal, Amit* (2001): Modern Information Retrieval: A Brief Overview. In: Bulletin of the IEEE Computer Society Technical Committee on Data Engineering, 24 (4): 35–42.
- Sinz, Elmar/Ulbrich-vom-Ende, Achim* (2015): CEUS - Systemarchitektur. URL: <http://www.ceushb.de/?content=architektur>, Abruf am 22.04.2015.
- Sinz, Elmar J./Ulbrich-vom-Ende, Achim* (2010): Architektur von Data-Warehouse-Systemen. In: Chamoni, Peter/Gluchowski, Peter (Hrsg.): Analytische Informationssysteme. 4. Aufl. Berlin, Heidelberg: Springer: 175–196.
- Siraj, Saba/Gupta, Ajay K./Badjugar, Rinku* (2012): Network Simulation Tools Survey. In: International Journal of Advanced Research in Computer and Communication Engineering, 1 (4): 201–210.
- Söllner, Fritz* (2015): Die Geschichte des ökonomischen Denkens. Berlin, Heidelberg: Springer.
- Sommerville, Ian* (2012): Software-Engineering, 9. Aufl. München et al.: Pearson.
- Souza, Bruno F. F./Batista, Maria C. M./Salgado, Ana C.* (2012): Semantic Loss in Query Reformulation in Dynamic Distributed Environments. In: Freire, Juliana/Suciu, Dan (Hrsg.): Proceedings of the 6th Alberto Mendelzon International Workshop on Foundations of Data Management. Aachen: RWTH Technical University of Aachen: 220–224.
- Stadtler, Hartmut* (2009): A framework for collaborative planning and state-of-the-art. In: OR Spectrum, 31 (1): 5–30.
- Statistisches Bundesamt* (2013): Statistisches Jahrbuch 2013. Wiesbaden: Statistisches Bundesamt.
- Stavrianos, Michael/Henderson, Don* (2006): Collaborative Business Intelligence: Integrating BI and KM. URL: <http://www.information-management.com/issues/20060601/1088414-1.html>, Abruf am 12.03.2014.
- Stegbauer, Christian/Jäckel, Michael* (2008): Social Software — Herausforderungen für die mediensoziologische Forschung. In: Stegbauer, Christian/Jäckel, Michael (Hrsg.): Social Software. Wiesbaden: Verlag für Sozialwissenschaften: 7–10.
- Stein, Friedrich* (2009): Projektmanagement für die Produktentwicklung, 3. Aufl. Renningen: Expert.
- Steininger, Katharina/Riedl, René/Roithmayr, Friedrich/Mertens, Peter* (2009): Moden und Trends in Wirtschaftsinformatik und Information Systems. In: WIRTSCHAFTSINFORMATIK, 51 (6): 478–495.
- Stichting CIS* (2015): Stichting Centraal Informatie System. URL: <http://www.stichtingcis.nl/>, Abruf am 11.05.2015.

- Stigler, George J.* (1957): Perfect Competition, Historically Contemplated. In: *Journal of Political Economy*, 65 (1): 1–17.
- Stonebraker, Michael* (1986): The Case for Shared Nothing. In: *Database Engineering*, 9 (1): 4–9.
- Strahringer, Susanne* (2013): Modell. URL: [www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/is-management/Systementwicklung/Hauptaktivitaeten-der-Systementwicklung/Problemanalyse-/konzeptuelle-modellierung-von-is/modell/index.html](http://www.enzyklopaedie-der-wirtschaftsinformatik.de/wi-enzyklopaedie/lexikon/is-management/Systementwicklung/Hauptaktivitaeten-der-Systementwicklung/Problemanalyse-/konzeptuelle-modellierung-von-is/modell/index.html), Abruf am 11.05.2015.
- Straube, Manfred* (1972): Zwischenbetriebliche Kooperation. Wiesbaden: Gabler.
- Stroh, Florian/Winter, Robert/Wortmann, Felix* (2011): Method Support of Information Requirements Analysis for Analytical Information Systems. In: *Business & Information Systems Engineering*, 3 (1): 33–43.
- Strong, Diane/Madnick, Stuart/Hartman, Amir/Peace, A. Graham/Teo, Thompson* (1994): Data Quality: A Critical Research Issue for the 1990s and Beyond. In: *Association for Information Systems (Hrsg.): Proceedings of the 15th International Conference on Information Systems (ICIS 1994)*: 500–501. URL: <http://aisel.aisnet.org/icis1994/18/>, Abruf am 08.12.2014.
- Surowiecki, James* (2007): Die Weisheit der Vielen, 2. Aufl. München: Goldmann.
- Swap, Walter C.* (1984): How Groups Make Decisions: A Social Psychological Perspective. In: *Swap, Walter C. (Hrsg.): Group decision making*. Beverly Hills, CA: Sage: 45–68.
- Swoboda, Bernhard* (2005): Erklärungsperspektiven grundlegender Theorien, Ansätze und Konzepte im Überblick. In: *Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke*. 2. Aufl. Wiesbaden: Gabler: 35–64.
- Sydow, Jörg* (1992): Strategische Netzwerke. Wiesbaden: Gabler.
- Sydow, Jörg/Duschek, Stephan* (2011): Management interorganisationaler Beziehungen. Stuttgart: Kohlhammer.
- Tayi, Giri K./Ballou, Donald P.* (1998): Examining Data Quality. In: *Communications of the ACM*, 41 (2): 54–57.
- Teich, Tobias/Fischer, Marco/Käschel, Joachim* (2002): Non-Hierarchical Production Networks. In: *Seuring, Stefan/Goldbach, Maria (Hrsg.): Cost Management in Supply Chains*. Heidelberg: Physica: 177–193.
- Teruel, Miguel A./Tardío, Roberto/Navarro, Elena/Maté, Alejandro/González, Pascual/Trujillo, Juan/Muñoz-Terol, Rafael* (2014): CSRML4BI: A Goal-Oriented Requirements Approach for Collaborative Business Intelligence. In: *Yu, Eric/Dobbie, Gillian/Jarke, Matthias/Purao, Sandeep (Hrsg.): Conceptual Modeling*. Berlin, Heidelberg: Springer: 423–430.
- The Data Warehousing Institute* (2013): Updated Style Intelligence Features collaborative BI. URL: <http://tdwi.org/articles/2013/02/12/style-intelligence-updated.aspx>, Abruf am 17.03.2015.
- The Law Dictionary* (2013): What is COLLABORATIVE BUSINESS INTELLIGENCE (CBI)? URL: <http://thelawdictionary.org/collaborative-business-intelligence-cbi/>, Abruf am 11.12.2013.

- Thierauf, Robert J.* (2001): Effective business intelligence systems. Westport, CT: Quorum Books.
- Thomae, Hans* (1960): Der Mensch in der Entscheidung. München: Barth.
- Thomson Reuters* (2015): Volumen der weltweiten M&A Deals von 2007 bis 2014 (in Milliarden US-Dollar). URL: <http://de.statista.com/statistik/daten/studie/153735/umfrage/volumen-der-fusionen-und-uebernahmen-weltweit/>, Abruf am 11.05.2015.
- TIOBE* (2015): TIOBE Index for March 2015. URL: <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>, Abruf am 23.03.2015.
- Torlone, Riccardo* (2008): Two approaches to the integration of heterogeneous data warehouses. In: Distributed and Parallel Databases, 23 (1): 69–97.
- Torrentfreak* (2014): Spotify Reminded of uTorrent Past After Branding Grooveshark ‘Pirates’. URL: <http://torrentfreak.com/spotify-reminded-of-utorrent-past-after-branding-grooveshark-pirates-141112/>, Abruf am 11.12.2014.
- Tresch, Markus* (1996): Middleware: Schlüsseltechnologie zur Entwicklung verteilter Informationssysteme. In: Informatik-Spektrum, 19 (5): 249–256.
- Trujillo, Juan/Maté, Alejandro* (2012): Business Intelligence 2.0: A General Overview. In: Aufaure, Marie-Aude/Zimányi, Esteban (Hrsg.): Business Intelligence. Berlin, Heidelberg: Springer: 98–116.
- Tseng, Frank S. C./Chen, Chia-Wei* (2005): Integrating heterogeneous data warehouses using XML technologies. In: Journal of Information Science, 31 (3): 209–229.
- Turban, Efraim/Sharda, Ramesh/Delen, Dursun* (2011): Decision support and business intelligence systems, 9. Aufl. Upper Saddle River, NJ: Pearson Prentice Hall.
- Turban, Ephraim/Liang, Ting-Peng/Wu, Shelly* (2011): A framework for adopting collaboration 2.0 tools for virtual group decision making. In: Group decision and negotiation, 20 (2): 137–154.
- Vaisman, Alejandro A./Minuto Espil, Mauricio/Paradela, Martín* (2009): P2P OLAP: Data model, implementation and case study. In: Information Systems, 34 (2): 231–257.
- Vera-Baquero, Alejandro/Colomo-Palacios, Ricardo/Molloy, Owen* (2013): Business Process Analytics Using a Big Data Approach. In: IT Professional, 15 (6): 29–35.
- Vetschera, Rudolf* (1991): Entscheidungsunterstützende Systeme für Gruppen. Heidelberg: Physica.
- VHB* (2011): VHB-JOURQUAL 2.1 Gesamtranking. URL: [http://vhbonline.org/uploads/media/Ranking\\_Gesamt\\_2.1.pdf](http://vhbonline.org/uploads/media/Ranking_Gesamt_2.1.pdf), Abruf am 11.05.2015.
- VHB* (2015): VHB-JOURQUAL 3 Teilrating Wirtschaftsinformatik. URL: <http://vhbonline.org/service/jourqual/vhb-jourqual-3/teilrating-wi/>, Abruf am 11.05.2015.



- Volkmann, Sabine* (2004): Unternehmensnahe Dienstleistungen im Licht der Statistik. In: Landesamt für Statistik Niedersachsen (Hrsg.): Statistische Monatshefte Niedersachsen. Hannover: Landesamt für Statistik Niedersachsen: 345–351.
- W3C* (2014): RDF - Semantic Web Standards. URL: <http://www.w3.org/RDF/>, Abruf am 11.05.2015.
- W3C* (2015): World Wide Web Consortium (W3C). URL: <http://www.w3.org>, Abruf am 11.05.2015.
- Walter, Rolf* (2007): Einführung in die Analysis. Berlin: de Gruyter.
- Wand, Yair/Wang, Richard Y.* (1996): Anchoring Data Quality Dimensions in Ontological Foundations. In: Communications of the ACM, 39 (11): 86–95.
- Wang, Richard Y./Strong, Diane M.* (1996): Beyond Accuracy: What Data Quality Means to Data Consumers. In: Journal of Management Information Systems, 12 (4): 5–33.
- Watson, Hugh J./Wixom, Barbara H.* (2007): The Current State of Business Intelligence. In: Computer, 40 (9): 96–99.
- Watts, Duncan J./Strogatz, Steven H.* (1998): Collective dynamics of 'small-world' networks. In: Nature, 393 (6684): 440–442.
- Watzlawick, Paul* (1981): Selbsterfüllende Prophezeiungen. In: Watzlawick, Paul (Hrsg.): Die erfundene Wirklichkeit. München, Zürich: Piper: 91–110.
- Weber, Andrea/Kolb, Martin* (2014): Collaborative BI - Gemeinsam mehr erreichen. In: is report Business Intelligence Guide 2014: 20–21.
- Weber, Karsten* (2004): Der wissenschaftstheoretische Status von Simulationen. In: Frank, Ulrich (Hrsg.): Wissenschaftstheorie in Ökonomie und Wirtschaftsinformatik. Wiesbaden: DUV: 191–210.
- Webster, Jane/Watson, Richard T.* (2002): Analyzing the Past to Prepare for the Future: Writing a Literature Review. In: MIS Quarterly, 26 (2): xiii–xxiii.
- Wehrle, Pascal/Miquel, Maryvonne/Tchounikine, Anne* (2005): A model for distributing and querying a data warehouse on a computing grid. In: IEEE (Hrsg.): Proceedings of 11th International Conference on Parallel and Distributed Systems (ICPADS). Los Alamitos, CA: IEEE: 203–209.
- Weitzel, Tim/Beimborn, Daniel/König, Wolfgang* (2003): Coordination in networks: An economic equilibrium analysis. In: Information Systems and e-Business Management, 1 (2): 189–211.
- Werner, Andrej/Hrach, Christian/Heyden, Karen/Alt, Rainer/Franczyk, Bogdan* (2010): Value Chain Cockpit auf Basis betrieblicher Anwendungssysteme. In: Schumann, Matthias/Kolbe, Lutz M./Breitner, Michael H./Frerichs, Arne (Hrsg.): Proceedings der Multikonferenz Wirtschaftsinformatik 2010 (MKWI 2010). Göttingen: Universitätsverlag Göttingen: 1583–1595.
- Wessler, Markus* (2012): Entscheidungstheorie. Wiesbaden: Gabler.
- Wilde, Thomas/Hess, Thomas* (2007): Forschungsmethoden der Wirtschaftsinformatik. In: WIRTSCHAFTSINFORMATIK, 49 (4): 280–287.
- Willimczik, Klaus* (2002): Erkenntnistheoretische Grundlagen erfahrungswissenschaftlicher Forschung. In: Singer, Roland/Willimczik, Klaus (Hrsg.): So-

- zialwissenschaftliche Forschungsmethoden in der Sportwissenschaft. Hamburg: Czwalina: 13–28.
- Winston, Wayne L. (2005): Operations research, 4. Aufl. Belmont, CA: Thomson; Brooks/Cole.
- Wissenschaftliche Kommission Wirtschaftsinformatik (2008): WI-Orientierungslisten. In: WIRTSCHAFTSINFORMATIK, 50 (2): 155–163.
- Wissenschaftliche Kommission Wirtschaftsinformatik (2011): Profil der Wirtschaftsinformatik. URL: [http://wi.vhbonline.org/fileadmin/Kommissionen/WK\\_WI/Profil\\_WI/Profil\\_WI\\_final\\_ds26.pdf](http://wi.vhbonline.org/fileadmin/Kommissionen/WK_WI/Profil_WI/Profil_WI_final_ds26.pdf), Abruf am 11.05.2015.
- Witten, Ian H./Frank, Eibe/Hall, Mark A. (2011): Data mining, 3. Aufl. Amsterdam et al.: Elsevier MK.
- Wittgenstein, Ludwig/Anscombe, Gertrude E. M./Hacker, Peter M. S./Schulte, Joachim. (2009): Philosophische Untersuchungen, 4. Aufl. Chichester, West Sussex: Wiley-Blackwell.
- Wolff, Heimfrid/Becher, Gerhard/Delpho, Holger/Kuhlmann, Stefan/Kuntze, Uwe/Stock, Johannes (1994): FuE-Kooperation von kleinen und mittleren Unternehmen. Heidelberg: Physica.
- Wölflle, Marco (2014): Mikroökonomik. Berlin, Heidelberg: Springer.
- Wooldridge, Jeffrey M. (2013): Introductory econometrics, 5. Aufl. Mason, OH: South-Western Cengage Learning.
- Wu, Sai/Jiang, Shouxu/Ooi, Beng C./Tan, Kian-Lee (2009): Distributed online aggregations. In: Proceedings of the VLDB Endowment, 2 (1): 443–454.
- Wurche, Sven (1994): Strategische Kooperation. Wiesbaden: DUV.
- Yatskevich, Mikalai/Giunchiglia, Fausto/McNeill, Fiona/Shvaiko, Pavel (2007): OpenKnowledge Deliverable 3.3: A methodology for ontology matching quality evaluation. Technical Report DIT-07-062. University of Trento, Trento.
- Yazici, Berna/Yolacan, Senay (2007): A comparison of various tests of normality. In: Journal of Statistical Computation and Simulation, 77 (2): 175–183.
- Zeng, Li/Li, Ling/Duan, Lian/Lu, Kevin/Shi, Zhongzhi/Wang, Maoguang/Wu, Wenjuan/Luo, Ping (2012): Distributed data mining: a survey. In: Information Technology and Management, 13 (4): 403–409.
- Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (2003): Kooperationen, Allianzen und Netzwerke - Grundlagen, "Metaanalyse" und Kurzaufsatz. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. Wiesbaden: Gabler: 3–34.
- Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.) (2005a): Kooperationen, Allianzen und Netzwerke, 2. Aufl. Wiesbaden: Gabler.
- Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (2005b): Kooperationen, Allianzen und Netzwerke - Grundlagen, "Metaanalyse" und Kurzaufsatz. In: Zentes, Joachim/Swoboda, Bernhard/Morschett, Dirk (Hrsg.): Kooperationen, Allianzen und Netzwerke. 2. Aufl. Wiesbaden: Gabler: 3–34.
- Zimmermann, Ekkart (1972): Das Experiment in den Sozialwissenschaften. Stuttgart: Teubner.



**Anhangsverzeichnis**

Anhang 1: Dokumentation zu CroCoSIM .....	267
Anhang 2: Modifizierter, iterativer Algorithmus zur Erzeugung von Zufallsnetzwerken .....	268
Anhang 3: Modifizierter Barabási-Albert-Algorithmus .....	269
Anhang 4: Beleg des Erhalts der Poisson-Verteilung von Knotengraden in Zufallsnetzwerken nach MRA.....	270
Anhang 5: Bestimmung Netzgrößen- und Gradanzahl-abhängiger Parameter für die geschätzte TNQ.....	271
Anhang 6: Betrachtung von NQ und TNQ ausgewählter Parameterkombinationen an sehr großen Netzwerken .....	272

## Anhang 1: Dokumentation zu CroCoSIM

Die Dokumentation zu CroCoSIM liegt als elektronische Ressource vor. In dem Hauptverzeichnis ‚CroCoSIM‘ befinden sich die folgenden Inhalte:

- Dokumentation: Eine vollständige Dokumentation des erstellten Quellcodes, generiert mit doxygen (<http://www.stack.nl/~dimitri/doxygen/>) und doxygraph (<https://code.google.com/p/doxygraph/>).
- Dimensionen.xlsx: Eine Excel-Datei, die die verwendeten Dimensionen zum Aufbau des Master-Würfels beinhaltet.
- license.txt: Die Hinweise zur Lizenzierung von CroCoSIM sowie die Verweise auf die verwendeten Komponenten Dritter mit entsprechenden Lizenzvermerken – insbesondere für die modifizierte GraphX-Bibliothek (<http://panthernet.ru/en/projects-en/graphx-en>).

Die hier beschriebenen Inhalte sind vollständig auf einem der Dissertation beigegebenen Datenträger vorhanden und stehen in der Version, wie sie zum Zeitpunkt der Abgabe oder Genehmigung der Veröffentlichung der Dissertation vorlagen, zum Download bereit unter: <http://udue.de/kaufmanndissanhang>

## Anhang 2: Modifizierter, iterativer Algorithmus zur Erzeugung von Zufallsnetzwerken

Dargestellt ist in C#-ähnlichem Pseudocode die grobe Struktur und Vorgehensweise des Algorithmus. Eine vollständige Dokumentation findet sich in Anhang 1.

```
MRA (peers, degree_target, degree_spread, corenet_size, PMQG, selfish,
    seed) {

    Initialisiere Zufallszahlengenerator(seed);
    Erstelle Zufallsliste rndpeers aller Peers in peers;
    Erstelle Netz net aus corenet_size der Peers in rndpeers;

    //Errechne Zielverbindungzahl der Peers
    while (degree_average < degree_target) {
        foreach (peer in rndpeers) {
            if (degree_average ≥ degree_target) {
                break;}
            if (rnd < degree_spread + epsilon) {
                //+ epsilon, da sonst Abbruch bei degree_spread==0
                peer_connections++;}}}

    //Füge Peers der Reihe nach ins Netz ein
    foreach (peer in rndpeers) {
        //Errechne PMQ
        foreach (target in net) {
            if (selfish) {
                peer_target_pmq = PMQG*MQ(peer, target)+(1-PMQG)*rnd;}
            else {
                peer_target_pmq = PMQG*MQ(target, peer)+(1-PMQG)*rnd;}}
        Sortiere Liste der Peers im Netz net_peers nach peer_target_pmq;
        Füge die top peer_connections Peers von net_peers als Nachbarn hinzu;
        Entferne peer aus rndpeers;}}
```

*Pseudocode des MRA*

*(Quelle: Eigene Darstellung)*

### Anhang 3: Modifizierter Barabási-Albert-Algorithmus

Dargestellt ist in C#-ähnlichem Pseudocode die grobe Struktur und Vorgehensweise des Algorithmus. Eine vollständige Dokumentation findet sich in Anhang 1.

```

BA (peers, m, mu, sorted, PMQG, selfish, seed) {

    Initialisiere Zufallszahlengenerator(seed);
    Erstelle Zufallsliste rndpeers aller Peers in peers;
    Erstelle Netz net aus m+1 der Peers in rndpeers;
    net_connectioncount = 0;

    //Füge Peers der Reihe nach ins Netz ein
    foreach (peer in rndpeers) {
        //Errechne PMQ
        foreach (target in net) {
            if (selfish) {
                peer_target_pmq = PMQG*MQ(peer, target)+(1-PMQG)*rnd;
            }
            else {
                peer_target_pmq = PMQG*MQ(target, peer)+(1-PMQG)*rnd;
            }
        }
        Sortiere Liste der Peers net_peers im Netz nach peer_target_pmq;
        peer_connectioncount = 0;
        i=0;
        //Erstelle m Verbindungen
        while (peer_connectioncount < m)
            if (sorted) {
                target = net_peers[i];
                iif (i==net_peers_count-1) ? i = 0 : i++;
            }
            else {
                target = net_peers[rnd*net_peers_count];
            }
        //Bestimme die Wahrscheinlichkeit einer Verbindung
        //Nutze mu zur Gewichtung der PMQ-Berücksichtigung
        //net_peers_maxd: maximale Gradanzahl eines Knoten in net_peers
        P_connect=(d(target)+mu*peer_target_pmq*net_peers_maxd) /
            (net_connectioncount * (2+2*mu))
        If (rnd < P_connect) {
            Füge target der Nachbarschaft von peer hinzu;
            peer_connectioncount++;
            net_connectioncount++;
        }
        Entferne peer aus rndpeers;
    }
}

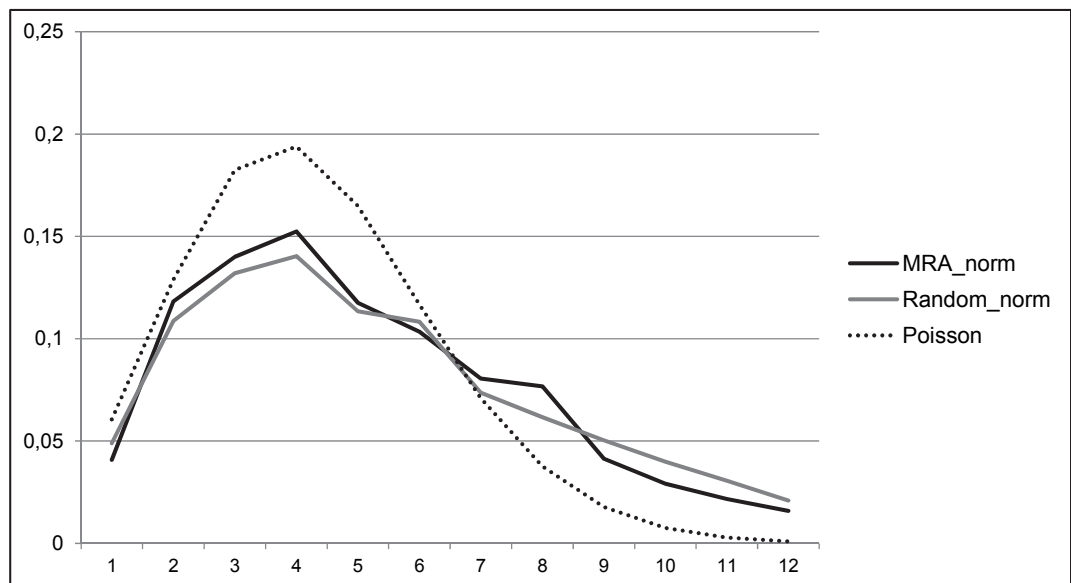
```

*Pseudocode des modifizierten BA-Algorithmus*

*(Quelle: Eigene Darstellung)*

#### Anhang 4: Beleg des Erhalts der Poisson-Verteilung von Knotengraden in Zufallsnetzwerken nach MRA

Die folgende Abbildung zeigt eine Verteilung der Grade aller Knoten in 7.260 Netzwerken, die mit Hilfe des modifizierten iterativen Algorithmus (MRA) zur Erstellung von Zufallsnetzwerken erzeugt wurden, im Vergleich zu einem unmodifizierten Algorithmus und der idealtypischen Poisson-Verteilung ( $\lambda = 4,25$ ). Deutlich erkennbar ist der grundsätzliche Erhalt einer Poisson-Verteilung (insbesondere im Vergleich zu den Potenzgesetz-proportionalen Verteilungen der skalenfreien Netzwerke) auch bei Berücksichtigung der Bedingungen zur Nachbarschaftswahl. Die im Vergleich zur Poisson-Verteilung erkennbare Stauchung der Kurve ist durch die Mitbetrachtung der Extremwerte *degree\_spread*  $\rightarrow 1$  erklärbar, die zu sehr hohen Abweichungen vom Zielknotengrad führen, da hier auch Knoten mit dem maximalen Grad 199 erzeugt werden.



Parameter	Beschreibung
Net size	200
Seed	1..3
Target degree	4..8
Core net size	1..5
Spread	0,0..1,0
PMQG	0,0..1,0
Selfish approach	True/False

*Darstellung der Knotengradverteilungen und verwendeten Parameter*

*(Quelle: Eigene Darstellung)*

## Anhang 5: Bestimmung Netzgrößen- und Gradanzahl-abhängiger Parameter für die geschätzte TNQ

Zur Ermittlung einer Gesamtfunktion zur Schätzung der TNQ wurden zunächst die sechzehn einzelnen Regressionsgeraden der Datenwertmengen errechnet, die sich bei einzelner Betrachtung von Netzgröße und Gradanzahl ergeben. Die folgenden Tabellen stellen die dabei ermittelten Werte der Steigungen der Geraden ( $m$ ) und der Achsenabschnitte ( $b$ ) sowie die Bestimmtheitsmaße der Regressionen ( $R^2$ ) dar:

	Größe	Gradanzahl					
		Unabhängig	Ø Abhängig	4	6	8	10
$m$	25	0,1165	0,2547	0,1715	0,2465	0,2945	0,3064
	50	0,0851	0,2041	0,1540	0,1910	0,2254	0,2460
	100	0,0838	0,1932	0,1408	0,1747	0,2083	0,2489
	200	0,0763	0,1779	0,1407	0,1608	0,1926	0,2175
$b$	25	0,0093	-0,0538	-0,0059	-0,0460	-0,0761	-0,0872
	50	0,0287	-0,0161	0,0148	-0,0072	-0,0287	-0,0434
	100	0,0394	0,0017	0,0305	0,0123	-0,0068	-0,0291
	200	0,0507	0,0199	0,0426	0,0290	0,0116	-0,0037
$R^2$	25	0,7893	0,7042	0,6972	0,7324	0,7109	0,6763
	50	0,6954	0,8538	0,8315	0,8649	0,8671	0,8517
	100	0,7302	0,9126	0,9256	0,9249	0,9142	0,8855
	200	0,7003	0,9232	0,9467	0,9294	0,9246	0,8921

$m$ ,  $b$  und  $R^2$  Gradanzahl-abhängiger und -unabhängiger Regressionsanalysen von TNQ und NQ

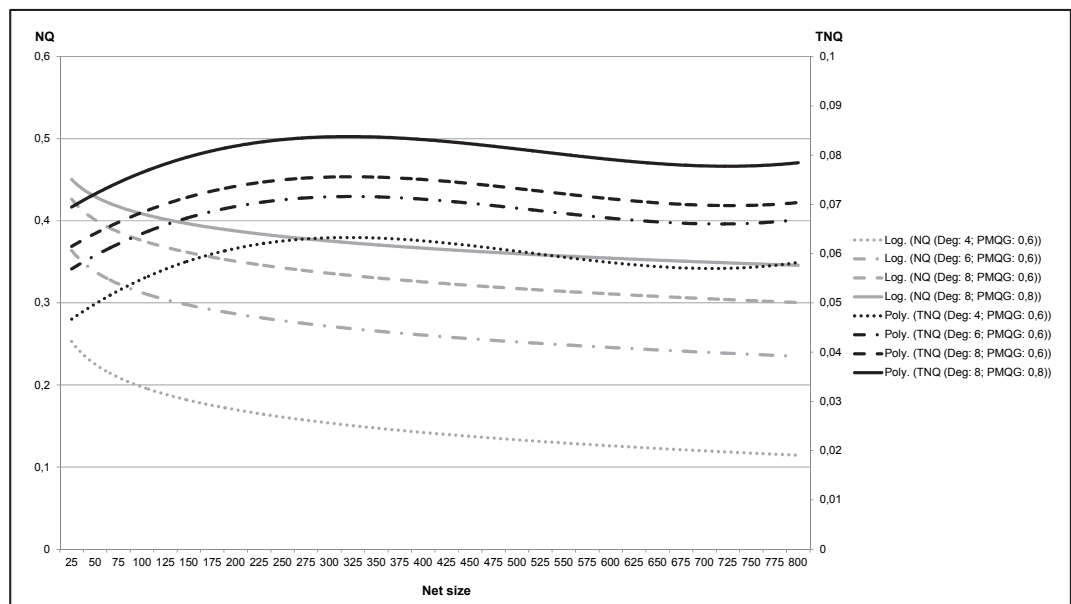
(Quelle: Eigene Darstellung)

Diese wurden im Anschluss wiederum genutzt, um die funktionale Abhängigkeit der Werte von  $m$  und  $b$  von der Netzwerkgröße und der Gradanzahl zu ermitteln, wobei diese zunächst durch den jeweils kleinsten verfügbaren Wert ( $n: 25; d: 4$ ) geteilt wurden und im Anschluss zur Basis zwei logarithmiert wurden.<sup>855</sup> Die dadurch ermittelten Werte für Steigung und Achsenabschnitt der zwei Regressionsgeraden für  $m$  und  $b$  (bzw. der Schätzungen  $\hat{m}$  und  $\hat{b}$ ) wurden im Anschluss verwendet, um die Steigung und den Achsenabschnitt der übergreifenden linearen Schätzfunktion ( $\widehat{TNQ}$ ) in Abhängigkeit von  $n$  und  $d$  zu setzen, sodass sich abschließend die in Abschnitt 6.4.4 ermittelte Formel ergibt.

<sup>855</sup> Vgl. dazu auch die Hinweise in Abschnitt 6.4.4.

## Anhang 6: Betrachtung von NQ und TNQ ausgewählter Parameterkombinationen an sehr großen Netzwerken

In der vorliegenden Arbeit wurde der Untersuchungsfokus auf Netzwerke ‚typischer‘ Größe in Bezug auf das vorgestellte Fallbeispiel gelegt. Sehr große Netzwerke mit deutlich mehr Teilnehmern ( $n \gg 200$ ) erscheinen für den gegebenen Fall aktuell nicht praxisrelevant. Zur Durchführung einer entsprechenden Simulation wären zudem erheblich aufwändigere Rechenstudien erforderlich.<sup>856</sup> Um zu verifizieren, ob die grundlegenden Betrachtungen zu NQ und TNQ weiterhin Gültigkeit aufweisen, wurden für ausgewählte Parameter sehr große Netzwerke im Vergleich zu kleinen und mittleren Netzwerken ( $25 \leq n \leq 800$ , 25er-Schritte) erstellt und ausgewertet. Die folgende Abbildung zeigt NQ und TNQ in logarithmisch (NQ,  $R^2 \approx 0,80$ ), bzw. polynomisch (TNQ, 3. Ordnung,  $R^2 \approx 0,65$ ) interpoliertem Verlauf. Gewählt wurde dabei jeweils die Annäherung mit dem höchsten Bestimmtheitsmaß. Es zeigt sich, dass die grundsätzlichen Zusammenhänge bestehen bleiben, die TNQ jedoch ab etwa  $n = 300$  zunächst wieder leicht zurückgeht, um danach erneut anzusteigen. Es bietet sich hier Untersuchungspotenzial hinsichtlich des mehrfach diskutierten Effekts der ‚kritischen Masse‘.<sup>857</sup>



*Darstellung interpolierter NQ und TNQ für kleine bis sehr große Netzwerke  
(Quelle: Eigene Darstellung)*

<sup>856</sup> Die Simulationszeit für 1.000 Netze mit 800 Teilnehmern bei Berechnung der TNQ beträgt annähernd 20 Stunden.

<sup>857</sup> Vgl. Abschnitte 6.4.3.2, 6.4.3.3, 7.1 und 7.2.